思科系列丛书

路由技术（IPv6 版）

张国清　编著

電子工業出版社
Publishing House of Electronics Industry
北京 • BEIJING

VISIT…

内 容 简 介

本书是一本全面、系统介绍 IPv6 路由协议、组播路由协议和 MBGP 的专业图书，具体内容包含 IPv6 BGP、RIPng、EIGRPv6、OSPFv3、IS-ISv6、IPv6 组播和 IPv6 过渡技术。为了更好地帮助读者掌握路由技术，本书配有光盘，以增强其实用性和可读性。

本书既可供期望获得思科职业认证资格的人员、网络工程师、系统工程师、网络管理员及高层技术管理人员参考阅读，也可作为思科网络技术学院相关专业的教材和教辅用书，还可作为员工的技术培训教材使用。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

路由技术：IPv6 版 / 张国清编著. —北京：电子工业出版社，2014.1
（思科系列丛书）

ISBN 978-7-121-21910-8

Ⅰ. ①路… Ⅱ. ①张… Ⅲ. ①计算机网络—路由选择 Ⅳ. ①TN915.05

中国版本图书馆 CIP 数据核字（2013）第 274530 号

策划编辑：宋 梅
责任编辑：宋 梅
印 刷：
装 订：
出版发行：电子工业出版社
北京市海淀区万寿路 173 信箱 邮编 100036
开 本：787×1 092 1/16 印张：16.5 字数：422 千字
印 次：2014 年 1 月第 1 次印刷
印 数：3 000 册 定价：49.00 元（含光盘 1 张）

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

前　言

路由协议是数据在互联网上有路径可用的唯一保证。网络技术工程师必须对路由协议非常熟悉。而路由协议又非仅有一个，掌握多个路由协议原理并不是一件容易的事。多数工程技术人员学习路由协议的途径是阅读思科的各种认证书籍，由简到繁，由易到难，逐步学习。思科的认证体系是分级别的，因此，路由协议的介绍也是在不同级别的认证教材中出现的。市场上缺少一本专门讨论各种路由协议的书籍，本书则正是填补这个空白之作。

本书在写作上突出两点：一是理论和实践相结合，在详细讲述路由协议的重点、要点、难点的同时，也使用完整实验环境验证路由协议的运行状态。二是详细介绍路由器操作界面上的信息含义和路由协议的重要参数，比如，OSPF 链路状态数据库中的各项信息的含义。当读者理解了这些信息的含义后，在日后工作中，就能够非常熟练地应用和管理路由协议。本书还根据技术的本质和各要点间的逻辑关系组织其内容结构，使其脉络清晰、知识连贯，使读者对路由技术有一个更系统、更清晰、更完整的理解。

参考过 CCIE 认证考试的人员应该都阅读过一本“名著”——《TCP/IP 路由》。这是一本引进的外版书籍，该书在内容构思上贯穿了协议开发的思路，在较高的理论层面讲解路由协议，对大多数读者来说，偏理论，详简不当。如果对路由协议没有系统了解的话，阅读该书会显得吃力。本书可以作为《TCP/IP 路由》这本“名著”的铺垫和弥补，如果读者先读本书，再读《TCP/IP 路由》则会顺利得多。

本书宗旨是：以东方思维模式谋篇，适合中国读者之口味；理论与实践紧密结合，弥补外版书籍之不足。

欢迎广大读者对本书内容提出建议和意见。

作者微博：http://weibo.com/gooltsing

作者信箱：gooltsing@sina.com

作　者

2013 年 11 月于北京

目　录

第 1 章　IPv6 地址 ····· 1

1.1　IPv6 地址概述 ····· 1
1.1.1　IPv6 地址的表示方法 ····· 1
1.1.2　IPv6 前缀表示法 ····· 2
1.1.3　IPv6 地址类型 ····· 3
1.2　IPv6 报文头 ····· 8
1.2.1　报文头格式 ····· 8
1.2.2　扩展报文头格式 ····· 9
1.3　本章小结 ····· 10

第 2 章　IPv6 基础 ····· 11

2.1　ICMPv6 ····· 11
2.2　ICMPv6 邻居重定向 ····· 12
2.3　邻居发现协议 ····· 14
2.3.1　邻居发现 ····· 14
2.3.2　路由器宣告/请求 ····· 17
2.3.3　优选默认路由器 ····· 18
2.4　无状态自动配置地址 ····· 19
2.5　DHCPv6 ····· 20
2.5.1　消息类型 ····· 22
2.5.2　报文结构 ····· 23
2.5.3　状态代码 ····· 23
2.5.4　地址和端口号 ····· 24
2.5.5　启用 DHCP 服务 ····· 24
2.5.6　中继代理与服务器使用的消息格式 ····· 26
2.5.7　配置中继代理 ····· 27
2.5.8　无状态 DHCPv6 ····· 27
2.6　DHCPv6 前缀代表 ····· 28
2.6.1　部署方式 ····· 28
2.6.2　配置前缀代表 ····· 29
2.6.3　验证 PD ····· 31
2.7　IPv6 的移动特性 ····· 33
2.7.1　移动 IPv6 节点的构成部件 ····· 33
2.7.2　移动 IPv6 节点的工作过程 ····· 34
2.7.3　移动性扩展报文头 ····· 35
2.8　本章小结 ····· 36

第 3 章　支持 IPv6 的 RIP 协议……37
3.1　RIPng 消息格式……37
3.1.1　路由条目格式……38
3.1.2　下一跳属性值……38
3.1.3　请求消息……38
3.1.4　响应消息……39
3.2　配置 RIPng……39
3.2.1　配置命令解释……39
3.2.2　配置示例……40
3.2.3　查看运行状态的命令……41
3.3　本章小结……43
第 4 章　支持 IPv6 的 EIGRP 协议……44
4.1　配置 EIGRPv6 的命令……44
4.1.1　配置示例……45
4.1.2　验证……46
4.2　调整相关参数……50
4.3　配置路由汇总……51
4.4　配置路由认证……53
4.4.1　配置认证……53
4.4.2　认证示例……54
4.5　本章小结……55
第 5 章　支持 IPv6 的 OSPF……56
5.1　OSPFv3 与 OSPFv2……56
5.1.1　相似之处……56
5.1.2　不同之处……57
5.2　数据包结构……57
5.3　Hello 数据包……59
5.4　IPv6 使用的 LSA 类型……60
5.4.1　LSA 的产生……60
5.4.2　LSA 类型……60
5.5　配置 OSPFv3……62
5.5.1　基本配置命令……62
5.5.2　检查 OSPFv3 的运行状态……68
5.5.3　NBMA 配置示例……78
5.5.4　路由总结……82
5.5.5　配置 OSPFv3 认证……83

5.5.6 配置优雅重启特性……84
5.6 本章小结……85
第 6 章 支持 IPv6 的 IS-IS 协议……86
6.1 为支持 IPv6 扩展的 TLV……86
6.1.1 IPv6 可达性 TLV……86
6.1.2 IPv6 接口地址 TLV……87
6.1.3 网络层协议 ID TLV……87
6.1.4 多拓扑 TLV……87
6.1.5 MT IS TLV……88
6.1.6 MT IPv4 可达前缀 TLV……88
6.1.7 MT IPv6 可达前缀 TLV……89
6.2 单拓扑模式和多拓扑模式……89
6.3 配置单拓扑单协议 IS-ISv6……90
6.3.1 基本配置……91
6.3.2 检查运行状态……96
6.4 配置单拓扑多协议 IS-ISv6……105
6.4.1 理解单拓扑模式特性……105
6.4.2 配置单拓扑多协议 IS-ISv6……106
6.4.3 检查运行状态……111
6.5 配置多拓扑 IS-ISv6……118
6.5.1 拓扑设计……118
6.5.2 迁移至多拓扑模式……118
6.6 本章小结……128
第 7 章 IPv6 BGP 序幕——BGPv4……129
7.1 BGP 与 AS……129
7.2 使用 BGP 时的考虑……130
7.3 BGP 概述……130
7.4 BGP 消息类型……131
7.5 BGP 邻接状态……134
7.6 路径属性……135
7.6.1 属性分类……135
7.6.2 BGP 属性……135
7.7 BGP 同步规则……141
7.8 路由选择……142
7.8.1 不予考虑的路由……142
7.8.2 选择最佳路由……143
7.8.3 实现多路径支持……144
7.9 BGP 路由反射……144

7.9.1 Split Horizon……145
7.9.2 路由反射器……146
7.9.3 路由反射器术语……146
7.9.4 路由反射器的规划……146
7.9.5 路由反射规则……147
7.9.6 路由反射器设计技巧……148
7.10 配置 BGP 路由协议……149
7.10.1 基本配置命令……149
7.10.2 环回接口与 BGP 会话……149
7.10.3 通告路由……150
7.10.4 BGP 配置示例……151
7.10.5 更改下一跳属性值……158
7.10.6 BGP 同步分析……159
7.10.7 配置路由反射器……173
7.10.8 通告聚合路由……174
7.10.9 配置对等体组……178
7.10.10 应用团体属性……180
7.10.11 其他命令……181
7.11 多宿主……182
7.12 BGP 案例……183
7.13 本章小结……192
第 8 章 多协议 BGP 之 IPv6 BGP……194
8.1 BGP-4 的多协议扩展……194
8.1.1 多协议可达 NLRI……194
8.1.2 多协议不可达 NLRI……195
8.1.3 NLRI 的结构……195
8.1.4 多协议特性通告……196
8.2 配置支持 IPv6 单播的 MBGP……196
8.2.1 基本配置命令……196
8.2.2 检查运行状态……199
8.3 把 IPv6 路由重分布到 mBGP……200
8.4 IPv6 邻居交换 IPv4 路由……205
8.5 IPv4 邻居交换 IPv6 路由……211
8.6 示例：调整本地优先值……214
8.7 本章小结……220
第 9 章 IPv6 的过渡技术……221
9.1 隧道的实现模式……221
9.2 手动隧道……222

9.3 自动 6to4 隧道 ····228
9.4 GRE over IP 隧道 ····232
9.5 自动 ISATAP 隧道 ····236
9.5.1 ISATAP 地址 ····236
9.5.2 配置 ISATAP 隧道 ····236
9.6 兼容 IPv4 地址隧道 ····241
9.7 6RD 隧道 ····244
9.7.1 6RD 的主要构成部件及术语 ····244
9.7.2 隧道终结点的计算 ····245
9.7.3 配置 6RD ····247
9.8 本章小结 ····252
参考文献 ····254

第 1 章　IPv6 地址

1.1　IPv6 地址概述

尽管科学家们开发了许多技术来节约当前使用的 IP 地址空间，但仍然不能满足互联网飞速发展对地址的需求，迫切需要开发新一代 IP 地址，1995 年，IETF 提出了下一代 IP 地址的方案，IPng——Internet Protocol Next Generation，标准草案编号是 RFC1752。IANA 为其指定了版本号为 6，称其为 IPv6。该地址长度是当前 IPv4 地址长度的 4 倍，即 128 比特长，可提供 2^{128} 个地址，足以满足互联网未来对地址的需要。

IPv6 地址是不同于 IPv4 的新一代地址体系，并非是 IPv4 的简单替换，相关的硬件和软件都要做相应的变化后才能使用它。因此，整个互联网不会立即更改为 IPv6 地址，IPv4 和 IPv6 需要共存一段时间。目前，已经建立了许多基于 IPv6 的实验床（Testbed），供全球的用户测试使用，例如 6bone。

IPv6 除了为互联网提供无限的地址空间外，还具有 IPv4 不具备的许多特性，例如：

- 安全性；
- 为 Internet 和 Intranet 提供统一的地址方案；
- 更方便地使用物理地址；
- QoS；
- 自动配置地址；
- 移动性；
- 比 IPv4 具有更多的层次（Hierarchy）结构（IPv4 只有网络和主机两层结构）。

提示：请读者查阅 IETF 网站和相关 RFC 文档了解以上及其他特性的详细描述。关于 IPv6 的正式、最新 RFC 是 2460。

1.1.1　IPv6 地址的表示方法

128 比特的 IPv6 地址非常长，书写起来比较麻烦，为了书写表达方便，使用十六进制（Hexadecimal）表示，并且每 4 个十六进制数字之间用双小圆点（：）隔开，共用 32 位十六进制数字，表达方法有以下 3 种。

1. 优选（Preferred）表达形式

例如：

① FED9 : BA98 : 7654 : 3210 : FEDC : BA98 : 7654 : 3210；

② 1080 : 0000 : 0000 : 0000 : 0008 : 0800 : 200C : 417A。

2. 压缩（Compressed）表达形式

由于地址中经常包含连续的“0”字符串、前导“0”字符，为了书写方便，可以使用压缩形式。在压缩地址格式中，前导“0”可以省略（结尾“0”绝对不能省略）；如果包含连续的“0”字符串，可以使用两个双小圆点（::）表示一个或多个为“0”的 16 比特组。

“::”在地址中只能出现一次，可以压缩包含在地址的开头、中间和结尾的为“0”的 16 比特组。例如：

① 1080 : 0000 : 0000 : 0000 : 0008 : 0800 : 200C : 417A，可以表示为

1080 : 0 : 0 : 0 : 8 : 800 : 200C : 417A　进一步可以表示为

1080 : : 8 : 800 : 200C : 417A

② FF01 : 0000 : 0000 : 0000 : 0000 : 0000 : 0000 : 0101　可以表示为

FF01 : : 101

③ 0000 : 0000 : 0000 : 0000 : 0000 : 0000 : 0000 : 0001　　可以表示为

: : 1

④ 0 : 0 : 0 : 0 : 0 : 0 : 0 : 0　可以表示为

: :

3. IPv6 和 IPv4 混合表达形式

这种表达形式用在 IPv6 和 IPv4 混合环境中，地址中的前 6 组是十六进制数值，最低两组是原 IPv4 地址，用十进制表达。例如：

① : : 13 . 1 . 68 . 3　　IPv4 地址是 13 . 1 . 68 . 3

② : : FFFF : 129 . 144 . 52 . 38　　IPv4 地址是 129 . 144 . 52 . 38

1.1.2　IPv6 前缀表示法

在 IPv6 地址中已经没有了子网掩码的概念。在 IPv6 地址中引入了一个称为前缀（Prefix）的概念，用来表示地址的最左端有多少个连续比特构成（Comprise）前缀。和 IPv4 中 CIDR 的表示方法类似。例如，下面的形式都是合法的表示方法，表示由 60 比特构成前缀。

① 12AB:0000:0000:CD30:0000:0000:0000:0000/60

12AB::CD30:0:0:0:0/60

12AB:0:0:CD30::/60

② 12AB:0:0:CD30:123:4567:89AB:CDEF/60

下面的形式都是非法的表示格式：

12AB : 0 : 0 : CD3/60　　结尾“0”不能省略。

12AB : : CD30 : : /60　　在地址中只能使用一次“::”。

1.1.3　IPv6 地址类型

IPv6 地址类型由地址中的高位比特（High-order bit）确定，这些比特称为格式前缀（Format Prefix，FP）。IPv6 的地址类型有：

- 全球单点传送地址 / 单播地址（Global Unicast Address）；
- 任意点传送地址 / 任意播地址（Anycast Address）；
- 本链路单点传送地址 / 本链路单播地址（Link-local Unicast Address）；
- 本站点单点传送地址 / 本站点单播地址（Site-local Unicast Address）；
- 环回（回送）地址（Loopback Address）；
- 多点传送地址 / 组播地址（Multicast Address）；
- 未指定地址（Unspecified Address）；

除了以上地址外，还有一些其他用途的地址，请读者自行查阅资料。

在讨论这些地址之前，需要学习如下几个术语。

- 节点（Node）：IPv6 网络中的设备；
- 链路（Link）：媒介或者通信设备，通过此设备或媒介接点可以在数据链路层进行通信，例如，Ethernet、PPP、Frame Relay 等；
- 链路层接口标志符（Data-link Identifier）：用来标志链路接口的数据链路层地址，例如，Ethernet 中使用 MAC 地址表示；
- 接口（Interface）：节点和链路的连接点，一个节点可能有多个接口，例如，路由器。
- 邻居（Neighbors）：和同一条链路相连的节点；
- 地址（Address）：接口的 IPv6 层标志，或者一系列接口的 IPv6 层标志，例如，组播地址代表的接口。

1. 全球单点传送地址

单点传送地址用来标志单一的接口。如果数据包的目的地是一个单点传送地址，那么它仅被该地址标志的接口所接收。单点传送地址有许多种，包括全球单点传送地址、本链路单点传送地址、本站点单点传送地址以及其他特殊用途的单点传送地址，如嵌套 IPv4 地址的 IPv6 单点传送地址。

接入 Internet 的 IPv6 节点需要使用被称为可聚合全球单点传送地址（Aggregatable Global Unicast Address），类似于 IPv4 中的公有地址。可聚合全球单点传送地址依靠分层体系（Hierarchy）使路由表变得容易管理，地址格式如图 1-1 所示。

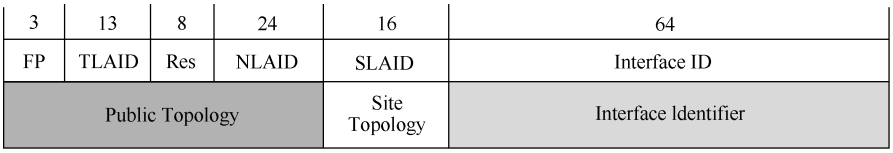

图 1-1　可聚合全球单点传送地址格式

总的来看，可聚合全球单点传送地址的结构可分公共拓扑部分、站点拓扑部分和接口标志号部分三部分。

公共拓扑部分代表提供 Internet 连接的服务提供商的集合，又分为以下几个部分。

- FP：格式前缀，可聚合全球单点传送地址的 FP=001；
- TLA ID：顶级聚合标志符（Top-Level Aggregation IDentifer），分配给提供公共传输拓扑的组织，例如，IANA、RIPE-NCC、INTERNIC 和 APNIC；
- Res：保留的 8 比特，为未来网络的扩展使用，当前值为“0”；
- NLA ID：下一级聚合标志（Next-Level Aggregation IDentifer），分配给 ISP 使用；
- SLA ID：站点层次聚合标志（Site-Level Aggregation IDentifer），公司 / 组织可以使用这些比特建立本地地址层次，用以标志不同的子网；
- Interface ID：在 IPv6 层上标志链路上的接口，类似 IPv4 地址中的主机部分，不过 IPv6 的接口标志起源于 IEEE EUI-64 格式。

如果是 LAN 接口，需要把 48 比特的 MAC 地址加 16 比特，形成 64 比特。转换方法是在原 48 比特的 MAC 地址中间（即在 OUI 部分和制造商为网卡分配的编号部分之间）插入 0xFF 和 0xFE，共 16 比特，并把全局 / 本地位取反。例如，MAC 地址。

c c c c c c 0 g c c c c c c c c c c c c c c c c x x x x x x x x x x x x x x x x x x x x x x x x

- “c”代表由 IEEE 分配给厂商的组织唯一标志符（OUI）；
- “0”代表全局 / 本地位；
- “g”代表个体 / 组（Individual/group）位；
- “x”代表网卡编号位。

转换成 IPv6 接口标志符的形式是：

c c c c c c 1 g c c c c c c c c c c c c c c c c 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 x x x x x x x x x x x x x x x x x x x x x x x x

如果把一个 EUI-64 地址转化为 IPv6 接口标志符，只需要把全局 / 本地位取反即可。例如，EUI-64 地址。

c c c c c c 0 g c c c c c c c c c c c c c c c c x x x x x x x x x x x x x x x x x x x x x x x x

转换成 IPv6 接口标志符的形式是：

c c c c c c 1 g c c c c c c c c c c c c c c c c x x x x x x x x x x x x x x x x x x x x x x x x

> **提示**：其他链路层 ID 与 IPv6 接口 ID 的转换方法请读者查阅 RFC3513。

配置 EUI-64 格式地址的命令如下：

```
Router(config-if)# ipv6 address 2001:0db8:1:: /64 eui-64
```

2. 任意点传送地址

任意点传送地址（任意播）是 IPv6 引入的一个新型地址。任意播地址是一组典型的（Typically）不属于同一个节点的接口的标志符；目的地是任意播地址的数据包被发向任意播组中最近的接口。

任意播地址从单播地址空间中分配（Allocation），因此，任意播地址和单播地址在语法上没有区别。当把一个单播地址指派给多个接口时，这样就变成了任意播地址，指派了该地址的

节点必须明确地配置任意播地址。

任意播地址有一个用来确定拓扑区的前缀“p”，属于这个区的接口都属于这个任意播组。在由“p”确定的区内，任意播地址必须作为单独的路由条目（通常被称为主机路由）被维护在路由表中，在“p”定义的拓扑区之外可以被聚合。

一个预定义的任意播地址是子网路由器任意播（Subnet-router Anycast）地址，地址格式如图 1-2 所示。

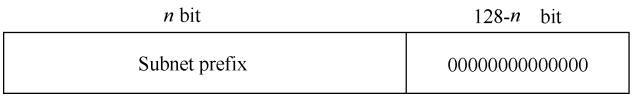

图 1-2　子网路由器任意播地址格式

子网前缀是标志（Identify）特定链路的前缀。与在该链路上的接口地址相比，它们的前缀是相同的，只是任意播地址的接口 ID 为“0”。目的地是子网路由器任意播地址的数据包只被发送到在该子网上的其中一台路由器（最近的那台）。要求所有在该子网有接口的路由器都支持该子网的任意播地址。子网路由器任意播地址被那些接点需要同一组路由器中的任意一台通信即可的应用程序使用。

任意播地址有以下约束条件：

- 任意播地址不能作为源地址使用；
- 任意播地址不能配置在 IPv6 的主机上使用，只能配置在 IPv6 的路由器上使用。

例如，在某路由器接口上配置了 IPv6 地址后，又配置子网路由器任意播地址和其他一个任意播地址：

```
Router(config-if)# ipv6 address 2001:0db8:1::1/64
Router(config-if)# ipv6 address 2002:0db8:6301::/128 anycast
Router(config-if)# 2001:DB8:1:1:FFFF:FFFF:FFFF:FFFE/64 anycast
```

3. 本链路地址

本链路地址仅用在单一链路上、为在单一链路上寻址的目的使用，例如，自动地址配置、邻居发现或者链路上不存在路由器（如只有几台 PC 互连的小型 LAN）。路由器不能把具有本链路源地址或目标地址的数据包转发到起它链路上。本链路地址的格式如图 1-3 所示。

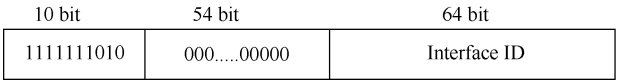

图 1-3　本链路地址格式

【例】　假设为一台 MAC 地址为 08-00-02-ab-12-34 的 PC 配置本链路地址，应该为

FE80 : 0000 : 0000 : 0000 : 0A00 : 02FF : FEAB : 1234

设备可以按照上述方法，自动为接口生成本链路地址，但也可以人为设置本链路地址。在路由器接口上配置本链路地址的命令如下：

```
Router(config-if)# ipv6 address ipv6-prefix / prefix-length link-local
```

例如：

```
DR(config-if)# ipv6 address FE80::260:3EFF:FE11:6770 link-local
```

4. 本站点地址

本站点地址用于在站点内寻址使用，即没有与全球 Internet 相连的网络，地址格式如图 1-4 所示。

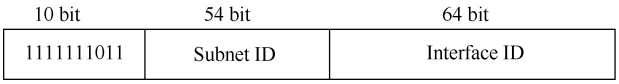

图 1-4　本站点地址格式

本站点地址以 FEC0::/10 开头，从地址结构格式上看，尽管它的子网 ID 有 54 比特长，当与 Internet 连接时仍然可以使用该地址。

路由器不能把具有本站点源地址和目标地址的数据包转发到站点以外。

5. 环回地址

单播地址为 ::1 的地址被定义为环回地址。节点使用该地址为自己发送数据包，类似 IPv4 中的 127.0.0.1，该地址不能指派给任何物理接口使用，并且具有本链路范围的作用域。

环回地址不能在 IPv6 数据包中做源地址使用，也不能被转发到节点之外，更不能被 IPv6 路由器转发。

6. 多点传送地址

在 IPv6 地址中没有广播的概念，广播地址被组播地址，即多点传送地址所替代。组播地址是标志一组接口的地址，典型地，这些接口属于不同的节点。发向组播地址的数据包被属于同一个组播组的所有接口接收。

多点传送地址的格式如图 1-5 所示。

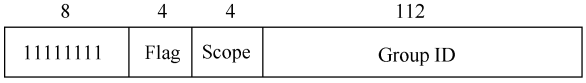

图 1-5　组播地址格式

各项参数解释如下：

- 组播地址的 FP=FF。
- 标志（Flag）：4 比特。前 3 比特保留，值为“0”。最后比特（RFC2373 把该比特命名为 T 比特）如果是“0”，代表永久分配的组播地址（IANA 指派的知名组播地址）；如果是“1”代表非永久性地址。RFC3306 在 T 比特前又增加了 P 比特的定义。如果 P 为 0，代表是一个普通组播；如果为 1，代表是一个包含数据源所在网络的单播前缀的组播。
- 作用域（Scope）：4 比特。定义组播影响的范围。
  - 1= 本接口（Interface-local）范围；
  - 2= 本链路（Link-local）范围；
  - 5= 本站点（Site-local）范围；
  - 8= 本组织（Organization-local）范围；
  - E= 全球（Global）范围；

- 0 和 F 保留，其他值未分配。

- 组 ID（Group ID）：112 比特，定义不同的组播组。

下面是一些预先定义的知名（Well-known）组播地址：

FF01:0:0:0:0:0:0:1　代表本接口范围内的所有 IPv6 节点；

FF01:0:0:0:0:0:0:2　代表本接口范围内的所有 IPv6 路由器；

FF02:0:0:0:0:0:0:1　代表本链路范围内的所有 IPv6 节点；

FF02:0:0:0:0:0:0:2　代表本链路范围内的所有 IPv6 路由器；

FF05:0:0:0:0:0:0:2　代表本站点范围内的所有 IPv6 路由器；

FF02:0:0:0:0:1:FF00:0000～FF02:0:0:0:0:1:FFFF:FFFF　用来生成被请求节点（Solicited-node）的组播地址，把被请求节点的单点传送地址或任意播地址的最后 24 比特写入到预留的空间中而形成。

下面是一些保留的组播地址：

FF00:0:0:0:0:0:0:0

FF01:0:0:0:0:0:0:0

FF02:0:0:0:0:0:0:0

FF03:0:0:0:0:0:0:0

FF04:0:0:0:0:0:0:0

FF05:0:0:0:0:0:0:0

FF06:0:0:0:0:0:0:0

FF07:0:0:0:0:0:0:0

FF08:0:0:0:0:0:0:0

FF09:0:0:0:0:0:0:0

FF0A:0:0:0:0:0:0:0

FF0B:0:0:0:0:0:0:0

FF0C:0:0:0:0:0:0:0

FF0D:0:0:0:0:0:0:0

FF0E:0:0:0:0:0:0:0

FF0F:0:0:0:0:0:0:0

7. 兼容 IPv4 的 IPv6 地址

该地址是内嵌 IPv4 地址的 IPv6 单播地址（IPv4-compatiable IPv6 Address），它的格式是最高 96 比特是 0，最后 32 比特是 IPv4 地址，如图 1-6 所示。例如，把 IPv4 地址 192.168.30.1 转换成兼容 IPv4 的 IPv6 地址转换后的地址是：::C0A8:1E01（或者::192.168.30.1）。

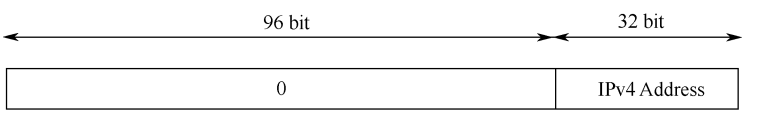

图 1-6　兼容 IPv4 的 IPv6 地址格式

在既配有 IPv4 地址和 IPv6 地址的接口上使用该地址比较方便。比如，在自动 IPv6 隧道模式下使用该地址可以方便建立穿越 IPv4 网络的 IPv6 隧道。此外，在移动 IPv6 环境下，移动节

点地址、基地地址、基地代理地址等也都使用兼容 IPv4 的 IPv6 地址。不过，这种地址目前不使用了。

8. 本地单播地址

本地单播地址（Unique Local Address）也是全球唯一的单播地址，只不过它用来在本网络内通信，不在全球范围内路由，类似于 IPv4 的私有地址，该地址格式如图 1-7 所示。

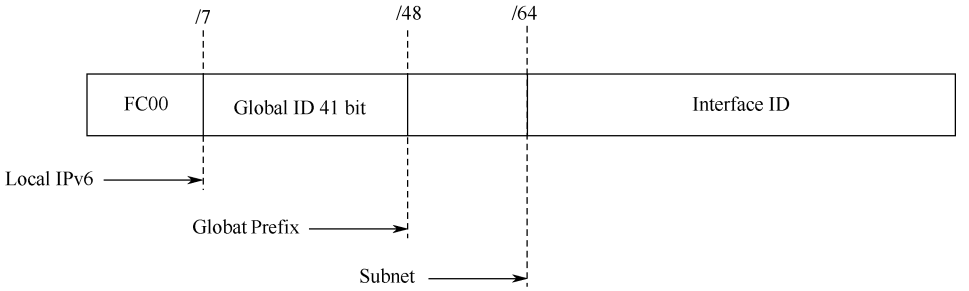

图 1-7　本地单播地址格式

本地单播地址的前缀以 FC00::/7 开头，紧随其后的 41 比特用来保证全球唯一，之后的 16 比特用来划分子网。

该地址有明确的前缀，所以很容易识别并在边界过滤此类路由，即便此类路由泄露到互联网上，因为有 41 比特的全球唯一性地址前缀，也不会造成地址冲突。

9. 未指定地址

地址 :: 被称为未指定地址，表示地址不存在。使用未指定地址的一个例子是当主机学习到自己的地址前，初始化时发送的数据包的源地址使用未指定地址。

该地址不能作为目的地地址使用。使用该地址作为源地址的数据包不被 IPv6 路由器转发。

1.2　IPv6 报文头

1.2.1　报文头格式

IPv6 的报文头格式如图 1-8 所示。和 IPv4 的报文头相比，没有了校验字段和与报文分段相关的字段。在 IPv6 中，网络层不在进行校验，而交由第二层和第四层负责校验。网络层设备也不在负责数据分段工作，由数据发送者负责数据的分段工作。

各字段含义解释如下。

- Version（版本号）：参数值为 6，代表 IPv6 报文头，4 比特长。
- Priority（优先级）：4 比特长。发送方可以使用该值为其发出的数据赋予优先级。
- Flow Label（流标签）：发送方可以用它标志属于同一个流的数据。
- Payload Length（载荷长度）：表示报文头后面携带的有效数据长度。
- Next Header（下一个报文头）：表示 IPv6 报文头后面的报文头类型，如 TCP、UDP

或其他扩展报文头。

- Hops（跳数限制）：表示报文所能经由的最大路由器数。
- Source Address（源地址）：发送数据方地址。
- Destination Address（目的地地址）：接收数据方地址。

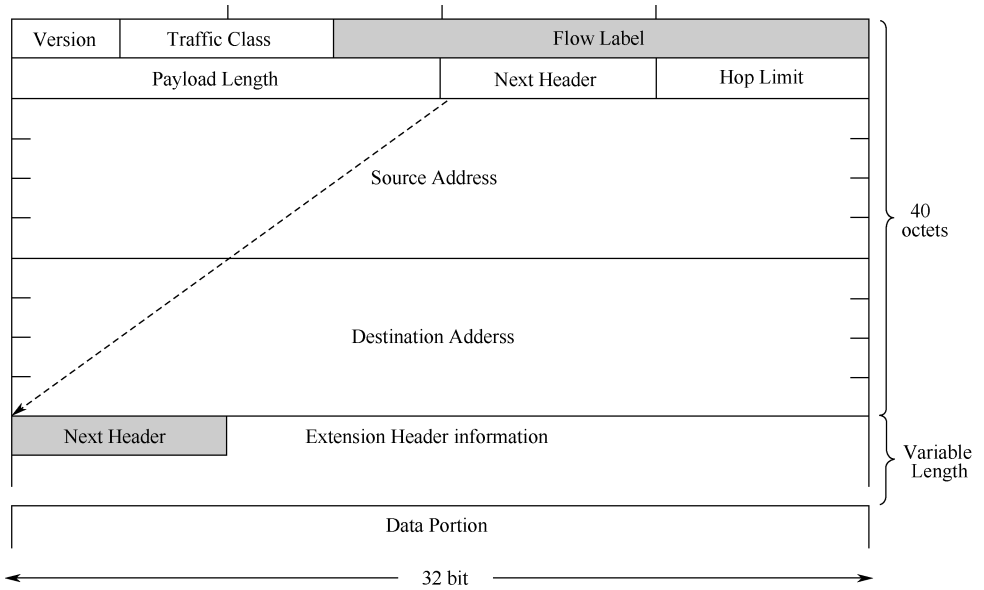

图 1-8　IPv6 报文头格式

1.2.2　扩展报文头格式

紧跟着 8 个基本字段后面是扩展报文头和数据部分。扩展报文头部分并不固定，如果存在，以 64 比特为单位附加。多个扩展报文头层层嵌套，内部扩展头由外层扩展头中的下一个报文头字段指示，如图 1-9 所示。如果数据包携带有上层数据，则最后一个下一个报文头字段用来指示传输层协议，比如 TCP 或 UDP。如果不携带上层数据，也就是没有下一个报文头，则下一个报文头的值为 59。

扩展报文头类型及代码如下所述。

- 逐跳扩展头（Hop-by-Hop Option）：代码是 0，用来携带一些可选信息，数据包所经由的路径上的所有路由器都必须处理该可选信息。可选信息的结构包括可选数据类型、可选数据长度和可选数据 3 部分。RFC2460 附录 B 介绍了如何定义可选数据结构。
- 目的地（Destination Option）扩展头：代码是 60，用来携带那些仅需要数据包目的地节点处理的可选信息。可选信息的结构包括可选数据类型、可选数据长度和可选数据 3 部分。
- 路由（Routing）扩展报文头：代码是 43，用来定义源路由。该扩展报文头中的数据部分列出了一系列数据包到达目的地必须经由的节点地址。
- 分段（Fragment）扩展报文头：代码是 44。当发送方必须对超出 MTU 的数据进行分段时使用该扩展报文头。扩展报文头在每一个数据片段中都存在。
- 认证头（Authentication Header）扩展报文头：代码时 51，封装 IPSec 数据。

- ESP 扩展报文头：代码是 50，封装 IPSec 数据。
- TCP：代码是 6，表示携带 TCP 数据。
- UDP：代码是 17，表示携带 UDP 数据。
- 移动性（Mobility）扩展报文头：代码是 135，移动节点等使用。

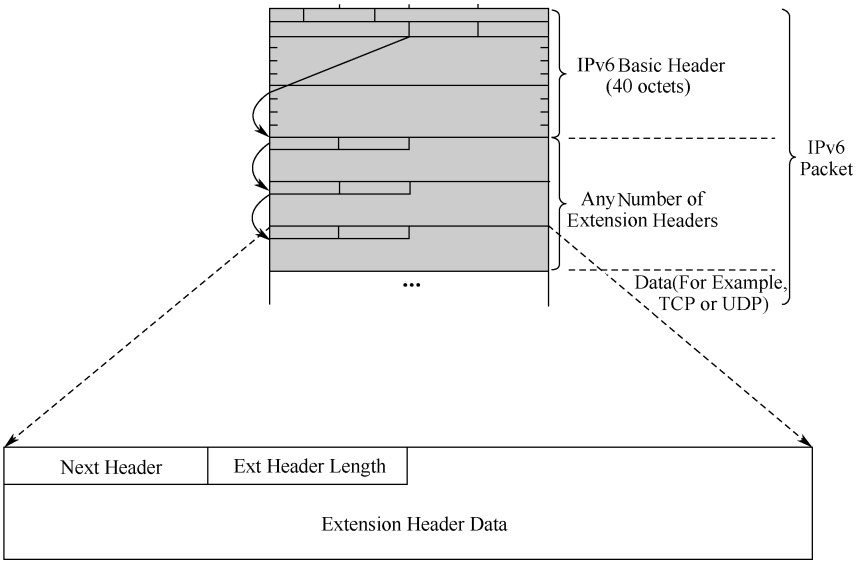

图 1-9　扩展报文头

1.3　本 章 小 结

本章介绍了新一代 IP 协议——IPv6，它是不同于 IPv4 地址的全新的 IP 地址体系，不仅地址空间是 IPv4 的 4 倍，而且还能提供新特性，如 QoS、定义流等。IPv6 地址主要有 3 种类型：单播地址、组播地址和任意播地址。

IPv6 既有基本的报文头结构，而且还有许多扩展报文头。扩展报文头提供了丰富的 IPv6 特性。常用的扩展头有：逐跳扩展头、路由扩展头、目的地扩展头、分段扩展头、认证扩展头、认证扩展头、ESP 扩展头、TCP 扩展头、UDP 扩展头、ICMP 扩展头和移动性扩展头。

第 2 章　IPv6 基础

2.1　ICMPv6

在 IPv6 协议体系中，ICMP 协议是一个重要协议，其功能与 IPv4 中的 ICMP 协议功能有许多相同的地方，但 ICMPv6 协议毕竟是一个新的协议，与 IPv4 中的 ICMP 协议相比有许多新的功能。比如，ICMPv6 同时完成 IPv4 中的 ICMP、IGMP 和 ARP 协议的功能。此外，在 IPv6 环境中，ICMPv6 还用来发现路径上支持的 MTU 值、发现邻居等。

ICMPv6 报文紧跟在 IPv6 的报文头后面，在 IPv6 报文头的下一个报文头字段中的数值是 58。ICMPv6 数据包在 IPv6 报文里类似传输层协议数据，它跟在可能存在的所有扩展报文头之后。在 ICMPv6 报文头中，类型和代码字段定义了携带的具体内容。

IPv6 节点使用 ICMPv6 报告处理数据包遇到的错误，也使用 ICMPv6 执行一些网络层功能，例如，故障诊断（ICMPv6 Ping）。因此，ICMPv6 消息可以分为两种类型：错误消息和信息消息。

错误消息的类型代码为 0～127；信息消息的类型代码为 128～255，ICMPv6 通用报文格式如图 2-1 所示。

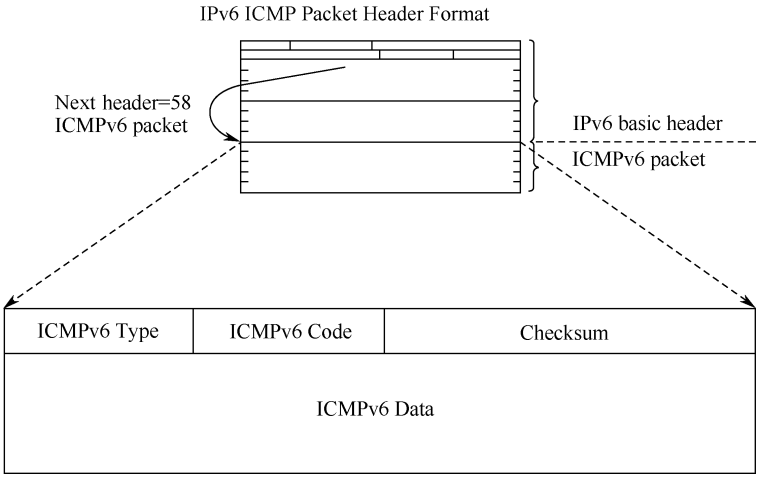

图 2-1　ICMPv6 报文格式

类型字段中的值表明了该信息的类型，也决定着后面所携带数据的格式。

代码字段中的值由信息类型决定，进一步划分了信息的子类。

校验和字段携带的是校验值。

错误消息包括：

- 类型 =1，目的不可达。

- Code=0，没有到达目的地的路由；
- Code=1，与目标的通信被管理性禁止（如被 ACL 禁止）；
- Code=2，超出源地址范围（如超出本链路地址作用范围）；
- Code=3，地址不可达（如地址解析失败）；
- Code=4，端口不可达；
- Code=5，源地址被入 / 出策略过滤；
- Code=6，目标路由被拒绝。

- 类型 =2，数据包过大。
  - Code=0。
- 类型 =3，超时。
  - Code=0，超出最大跳数限制；
  - Code=1，分段重组超时。
- 类型 =4，参数错误。
  - Code=0，报文头错误；
  - Code=1，不可识别的下一个报文头类型；
  - Code=2，不可识别的 IPv6 选项。
- 类型 =100，101 实验用。

信息消息包括：

- 类型 =128，回应请求（Echo Request）。
- 类型 =129，回应应答（Echo Reply）。
- 类型 =130，组成员查询。
- 类型 =131，组成员报告。
- 类型 =132，组成员减少。
- 类型 =133，路由器请求（Router Solicitation）。
- 类型 =134，路由器宣告（Router Advertisement）。
- 类型 =135，邻居请求（Neighbor Solicitation）。
- 类型 =136，邻居宣告（Neighbor Advertisement）。
- 类型 =137，重定向（Redirect）。
- 类型 =200，201 实验用。

提示：根据 RFC4443，类型代码 127 和 255 保留，用作未来的扩展。

2.2　ICMPv6 邻居重定向

在 ICMPv6 报文头中，类型字段的值如果是 137，代表这是一个邻居重定向消息。路由器使用该消息通知主机更好的下一跳节点，也就是更好的网关地址。如图 2-2 所示，主机 H 的网关是路由器 A，当主机把数据发给路由器 A 后，路由器 A 通过查阅路由表发现，到达目的地的下一跳是路由器 B，而且路由器 B 和主机 H 在同一个网段。如果主机 H 直接把数据发给路由器 B 则更好。因此，路由器 A 向主机 H 发送邻居重定向消息，通知 H 下次把数据发给路由

器 B，主机 H 后续的数据包都直接发送给路由器 B。

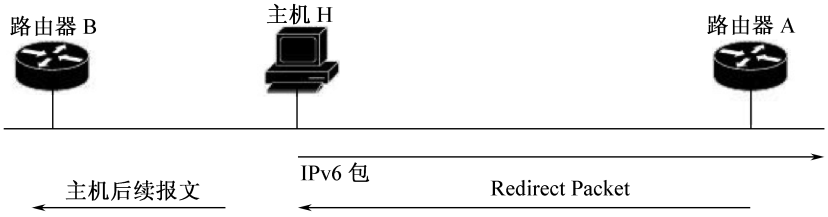

图 2-2　邻居重定向

邻居重定向报文是这样封装的：路由器 A 的本链路地址填充在源地址字段；主机 H 的本链路地址填充在目标地址字段；数据部分是路由器 B 的本链路地址。

为保证能正确封装邻居重定向数据包，路由器 A 必须能够确定链路上的每一个邻居的本链路地址。因此，如果是静态路由，下一跳地址必须使用邻居的本链路地址；如果是动态路由，也必须使用邻居设备的本链路地址交换路由信息。

在如下情况下，路由器发送第一个数据包后，应该发送邻居重定向消息：

- 目的地址不是组播地址；
- 目的地址不是本设备地址；
- 数据包将从其被接收的端口发送出去；
- 可以判断更好的出口节点和发送数据的节点在同一个链路上；
- 数据包的源地址是同链路上的某个邻居的全局地址或本链路地址。

路由器收到邻居重定向消息后不可以更新自己的路由表；主机不可以产生邻居重定向消息。

查看接口的命令可以显示接口上邻居重定向功能是否开启，例如：

```
Drouter# show ipv6 interface fastethernet 1/0
Ethernet0 is up, line protocol is up
  IPv6 is stalled, link-local address is FE80::1
  Global unicast address(es):
    2001:DB8:2000::1, subnet is 2001:DB8:2000::/64
    2001:DB8:3000::1, subnet is 2001:DB8:3000::/64
  Joined group address(es):
    FF02::1
    FF02::2
    FF02::1:FF00:1
  MTU is 1500 bytes
  ICMP error messages limited to one every 100 milliseconds
  ICMP redirects are enabled
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
  ND advertised reachable time is 0 milliseconds
  ND advertised retransmit interval is 0 milliseconds
  ND router advertisements are sent every 200 seconds
```

```
ND router advertisements live for 1800 seconds
Hosts use stateless autoconfig for addresses.
```

2.3　邻居发现协议

为 IPv6 开发的邻居发现协议（Neighbor Discovery Protocol，NDP），由 RFC2461 定义，用来帮助节点（主机和路由器）发现本链路上其他邻居的数据链路层地址。

主机使用邻居发现协议发现邻近的路由器，并把它作为自己的默认网关。节点使用邻居发现协议主动跟踪邻居是否可达，检测邻居数据链路层地址是否改变。当路由器或到达路由器的路径失效时，主机依靠该协议主动搜索可用的路由器或路径。概括起来，邻居发现协议解决的是在同一链路上的节点之间的交互问题，这些问题包括：

- 路由器发现。
- 前缀发现。
- 参数发现。
- IPv6 地址自动配置。
- 地址解析。
- 下一跳确定。
- 邻居不可达检测。
- 重复地址检测。
- 重定向。

邻居发现协议（NDP）使用以下 5 种类型的 ICMPv6 数据包工作。

- 路由器请求（133）：当接口启动后，主机发送该消息请求路由器立即产生路由器宣告消息。
- 路由器宣告（134）：路由器定期或接收到路由器请求信息后，使用该信息向链路上宣告它的存在。该数据包中携带有用来进行地址自动配置的前缀等信息。
- 邻居请求（135）：节点使用该信息确定邻居的数据链路层地址，或缓存的邻居的数据链路层地址是否可达。该信息也用来进行重复地址检测。
- 邻居宣告（136）：对邻居请求信息的回应信息。节点也可以发送未被请求的邻居宣告（Unsolicited Neighbor Advertisement）信息通告数据链路层地址的更改。
- 重定向：路由器用来通知主机更好的第一跳（网关）的信息。

综上所述，IPv6 的邻居发现协议综合了 IPv4 协议中的 ARP、ICMP 路由器发现和 ICMP 重定向功能，并具备在 IPv6 中工作的新功能。

2.3.1　邻居发现

在 IPv4 中，发送数据的节点已知目标节点的 IPv4 地址后，要想得到目标节点的数据链路层地址（这里指 MAC 地址），它唯一的做法是使用 ARP 解析。在 IPv6 中已经没有了 ARP 协议，这些功能包含在 ICMPv6 中。

当一个接口启动之后，节点就加入该接口上的所有节点组播组，即 FF02::1，同时，节点

也必须加入到指派给接口的每一个 IPv6 地址对应的被请求节点（Solicited-node）组播组中。被请求节点组播地址具有本链路作用范围，该地址是将接口 IPv6 地址的最后 24 比特复制到预定义的组播地址空间的最后 24 比特而构成的。预定义的被请求节点的组播地址空间是：FF02:0:0:0:0:1:FF00:0000～FF02:0:0:0:0:1:FFFF:FFFF。

例如，一个接口的 IPv6 地址是 2006:1010:0:1:20E:38FF:FEEE:39A0 / 64，那么，其被请求节点的组播地址是 FF02::1:FFEE:39A0。

下面以图 2-3 为例介绍 IPv6 中的节点是如何使用 ICMPv6 信息发现邻居的。

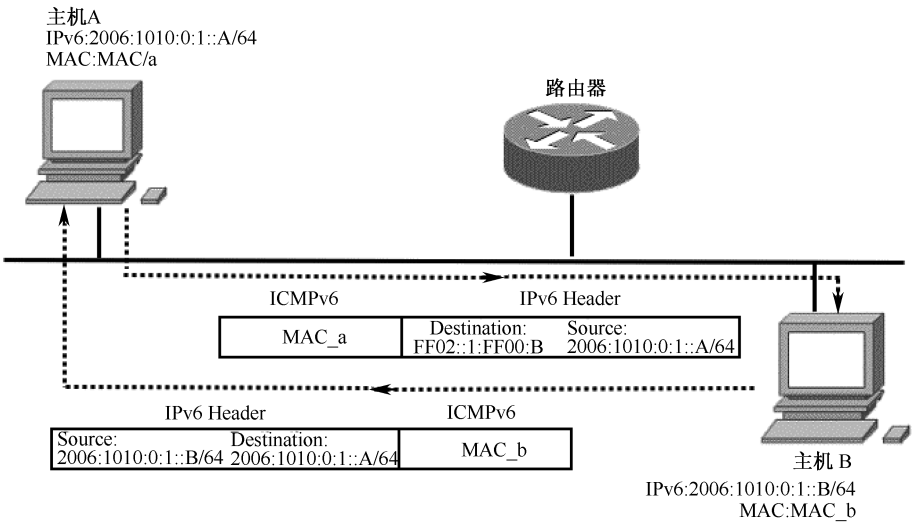

图 2-3　邻居发现过程

主机 A 无论是把数据转发给主机 B 还是转发给路由器，都需要获得目标节点的数据链路层地址。图 2-3 中描述的是主机 A 解析主机 B 的数据链路层地址的过程。

为了获得主机 B 的数据链路层地址，主机 A 向组播地址 FF02::1:FF00:B（主机 B 的被请求节点组播地址）发送一个邻居请求消息。邻居请求消息的 ICMPv6 报文头的构成如下：

- 类型= 135；
- 代码= 0；
- 在 ICMPv6 的数据字段里，填充主机 B 的 IPv6 地址和主机 A 的 MAC 地址 MAC_a。

这样，主机 B 收到该消息后就获得了主机 A 的数据链路层地址，并把它放置在自己的高速缓存区中。

主机 B 收到对自己的邻居请求消息后，使用邻居宣告消息应答。邻居宣告消息的报文结构和邻居请求消息的格式有所不同，如图 2-4 所示。

邻居宣告消息的 ICMPv6 报文头的构成如下：

- 类型= 136；
- 代码= 0；
- S 比特位= 1，表示该报文是对邻居请求消息的回应；
- R 比特表示是路由器的回应；
- O 比特表示需要更新（Override）缓存里的地址映射；

- 目标地址=主机 B 的地址；
- 选项=主机 B 的 MAC 地址。

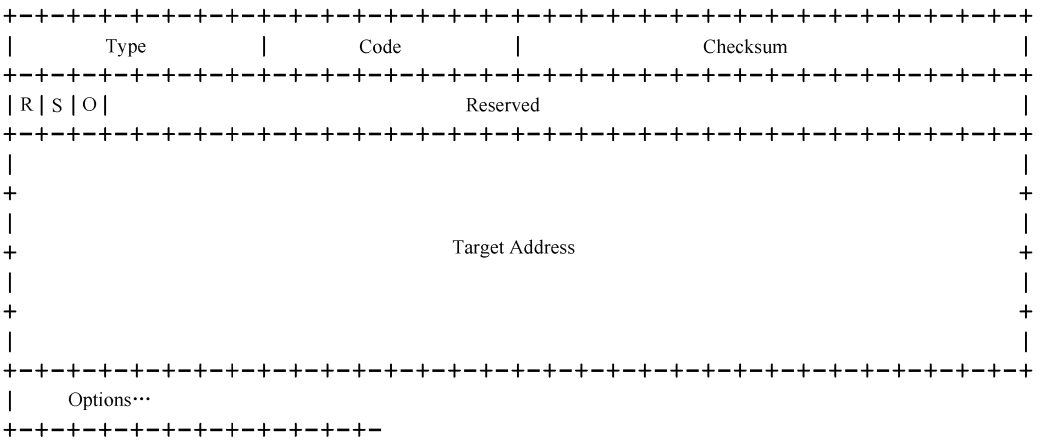

图 2-4　邻居宣告消息报文结构

主机 A 收到来自主机 B 的邻居宣告消息后，就获得了主机 B 的数据链路层地址，并把它放置在自己的高速缓存中。

邻居请求消息也用来检验邻居可达性，在这种情况下，目标地址是被测试邻居的单播地址。如果能收到邻居的确认消息或邻居宣告消息，表明邻居是可达的。

邻居宣告消息也用来通告链路地址变化消息，如果某个节点链路层地址变化了，它就使用邻居宣告消息把新的地址通告给其他邻居。这种邻居宣告消息称为非请求的邻居宣告（Unsolicited Neighbor Advertisements，NA），因为它不是为响应邻居请求消息而发送的，而是主动发送的。非请求邻居宣告消息的目标地址使用代表所有节点的组播地址。

关于邻居发现的配置命令如下所述。

① 定义缓存超时时间。

router(config-if)# **ipv6 nd cache expire** *expire-time-in-seconds*

② 邻居不可达性测试（Neighbor Unreachability Detection, NUD）次数。

router(config-if)# **ipv6 nd nud retry** *base interval max-attempts*

③ 从非请求邻居宣告信息中获取地址信息。

router(config-if)# **ipv6 nd na glean**

④ 设置接口下缓存的链路层地址的数量，

router(config-if)#i**pv6 nd cache interface-limit** *size*

或在全局模式下。

router(config)#i**pv6 nd cache interface-limit** *size*

⑤ 设置发送 ICMP 消息的频率。

router(config)#**ipv6 icmp error-interval** *milliseconds*

增大时间，降低发送各种类型的 ICMPv6 信息，可以缓解链路层的拥塞。

2.3.2　路由器宣告/请求

1. 路由器宣告

路由器从启用了 IPv6 的接口上定期发送路由器宣告（Router Advertisement，RA）消息。RA 使用的源地址是发送该消息的接口的本链路地址，目标地址是所有节点组播地址，便于让包括主机在内的其他设备发现本链路上的路由器。比如，主机收到 RA 后，可以获得网络前缀，从而为自己配置 IPv6 地址。

RA 使用 ICMPv6 报文，类型代码是 134。报文结构如图 2-5 所示。

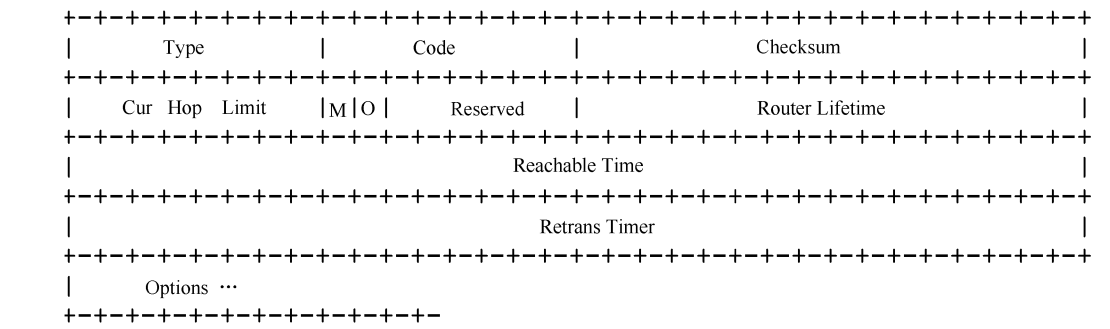

图 2-5　RA 报文结构

RA 的 ICMPv6 报文的主要字段解释如下。

- 类型= 134。
- 代码= 0。
- 跳数限制：默认应该是 IPv6 报文头中跳数字段的值，0 值代表未定义。
- M：如果该比特为 1，主机使用 DHCPv6 配置地址等参数，而不是使用无状态地址自动配置。
- O：如果该比特为 1，主机使用 DHCPv6 配置除地址外的其他参数。
- 路由器生命期：该参数仅与路由器是否是默认路由器有关。默认值是 18.2 小时，0 值代表该路由器不是一个默认路由器，并且不能列入默认路由器列表中。
- 可达时间：节点对路由器做 NUD 之后，就假设该路由器在这个时间段内都是可达的。
- 重传计时器：发送邻居请求消息和 NUD 的间隔时间。

常用的选项内容如下所述。

- 发送 RA 接口的链路层地址。
- MTU。
- 地址前缀：接口上设置的除本链路地址前缀的其他所有地址前缀，用来支持自动配置地址功能的实现。

路由器除了定期发送 RA 外，还使用 RA 响应路由器请求消息。此时，目的地址使用的是被响应者的单播地址。

如果禁止接口发送 RA，可以使用如下命令：

```
router(config-if)# ipv6 nd ra suppress
```

2. 路由器请求

为了让路由器立即发送 RA，而不是等到发送 RA 周期到才发送，主机发送路由器请求消息。主机在系统启动时发送路由器请求消息可以快速为自己自动配置 IPv6 地址。此时，IP 报文头的源地址是未定义地址（0:0:0:0:0:0:0:0），目的地址是代表本链路所有路由器的组播地址。

路由器请求消息的报文结构如图 2-6 所示。

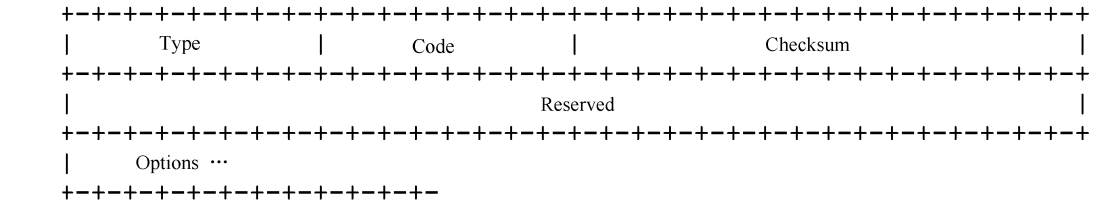

图 2-6　路由器请求报文

报文头构成如下：

- 类型= 133。
- 代码= 0。
- 保留字段= 0。
- 选项部分可以是发送者的链路层地址。如果源地址是未定义的 IPv6 地址，该部分应该为空。

2.3.3　优选默认路由器

主机是通过监听 RA 发现并选择默认路由器的，这种选择机制在某些情况下选出的默认路由器可能不是最理想的。比如，在同一条链路上有两台路由器，它们提供的路由是不等值的，主机应该选择提供最近路由的那台作为网关才理想。类似的其他情况还有，比如流量工程，减少重定向，避免错误部署的新设备成为默认网关等。

优选默认路由器特性（Default Router Preference，DRP）为路由器提供了一个基本的优先度量值，将路由器成为默认路由器的级别划分为低、中、高三个级别。该信息使用 RA 消息中未使用比特表示，主机和路由器皆可识别。如果主机不支持 DRP 特性，则忽略该比特即可；如果路由器不支持 DRP 而主机支持该特性，主机则默认该路由器的 DRP 为中级。

DRP 功能需要手工配置启用，命令如下：

```
Device(config-if)#ipv6 nd router-preference {high | medium | low}
```

可以查看接口下 DRP 设置的级别：

```
Device# show ipv6 interface gigabitethernet 0/1
GigabitEthernet0/1 is up, line protocol is up
  IPv6 is enabled, link-local address is FE80::130
  Description: Management network (dual stack)
  Global unicast address(es):
    FEC0:240:104:1000::130, subnet is FEC0:240:104:1000::/64
  Joined group address(es):
```

```
        FF02::1
        FF02::2
        FF02::1:FF00:130
    MTU is 1500 bytes
    ICMP error messages limited to one every 100 milliseconds
    ICMP redirects are enabled
    ND DAD is enabled, number of DAD attempts: 1
    ND reachable time is 30000 milliseconds
    ND advertised reachable time is 0 milliseconds
    ND advertised retransmit interval is 0 milliseconds
    ND router advertisements are sent every 200 seconds
    ND router advertisements live for 1800 seconds
    ND advertised default router preference is Low
    Hosts use stateless autoconfig for addresses
```

如果查看其他路由器通告的 DRP 级别，可以使用如下命令：

```
Device# show ipv6 routers
  Router FE80::169 on GigabitEthernet0/1, last update 0 min
    Hops 64, Lifetime 1800 sec, AddrFlag=0, OtherFlag=0, MTU=1500
    Preference=Medium
    Reachable time 0 msec, Retransmit time 0 msec
    Prefix FEC0:240:104:1000::/64 onlink autoconfig
      Valid lifetime 2592000, preferred lifetime 604800
```

2.4　无状态自动配置地址

所有 IPv6 节点的接口都能自动生成一个本链路地址，用来和本链路上的其他节点通信。IPv6 节点的接口也可以自动配置一个全球可路由单播地址，这样免去了手工配置地址或 DHCP 服务器。因为不需要任何服务器维护客户端的状态，所以称为无状态地址自动配置（Stateless Address Autoconfiguration）。

启用 IPv6 的路由器接口定期发送 RA 消息，或为响应主机发出的路由器请求消息而不定期发送 RA 消息。RA 消息中携带接口所配置的 IPv6 地址前缀，这样，主机把自己的接口标志附加在该地址前缀后即可生成自己的 IPv6 地址，这样便完成了无状态自动地址配置。自动配置的地址仍然需要考虑地址的唯一性，只要 RA 中通告的地址前缀是唯一的，不冲突的，主机自动生成的地址就是唯一的。系统一旦启动，主机就发送类型 133 的 ICMP 报文，路由器立即响应该报文，这样主机就马上获得的自己的地址。

这个特性也为主机更改地址提供了方便，让路由器同时通告新旧两个地址前缀，然后停止通告旧的地址前缀即可完成地址更改。

启用自动地址配置的命令很简单：

```
Device(config-if)# ipv6 address autoconfig
```

启用自动地址配置的接口状态如下：

```
Device# show ipv6 interface gigabitethernet 0/1
GigabitEthernet0/1 is up, line protocol is up
  IPv6 is enabled, link-local address is FE80::130
  Description: Management network (dual stack)
  Global unicast address(es):
    FEC0:240:104:1000::130, subnet is FEC0:240:104:1000::/64
  Joined group address(es):
    FF02::1
    FF02::2
    FF02::1:FF00:130
  MTU is 1500 bytes
  ICMP error messages limited to one every 100 milliseconds
  ICMP redirects are enabled
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
  ND advertised reachable time is 0 milliseconds
  ND advertised retransmit interval is 0 milliseconds
  ND router advertisements are sent every 200 seconds
  ND router advertisements live for 1800 seconds
  ND advertised default router preference is Low
  Hosts use stateless autoconfig for addresses
```

2.5　DHCPv6

为设备配置 IPv6 地址的方法有如下 5 种：

- 手工配置；
- 无状态自动配置；
- DHCPv6 有状态自动配置（Stateful DHCPv6）；
- 无状态 DHCPv6（Stateless DHCPv6）；
- DHCPv6 前缀代表（DHCPv6 Prefix Delegation，DHCPv6-PD）。

就 DHCPv6 有状态自动地址配置而言，IPv6 环境中的 DHCP 元素构成和 IPv4 网络环境中的 DHCP 元素构成相同，也由客户端、服务器和中继代理构成。如果客户端需要获得地址等配置参数，它向本网发送请求信息，在本网上的 DHCPv6 服务器则回应客户端需要的参数，如图 2-7 所示。

如果客户端和 DHCPv6 服务器不在同一个网段上，则由中继代理代为客户端转发请求信息，基本过程如图 2-8 所示。

客户端和服务器分别有一个独立的 ID 号，称为 DUID（DHCP Unique IDentifier）。在交换 DHCPv6 消息时，用来确定设备的身份。DUID 在 DHCPv6 消息的选项字段携带，该字段是变长字段，但不超过 128 字节。该选项字段不要求出现在所有 DHCPv6 消息中。目前，DUID 有以下 3 种类型：

- 链路层地址加时间（Link-layer address plus time，DUID-LLT）；
- 厂商基于企业代码指定的唯一 ID；
- 链路层地址。

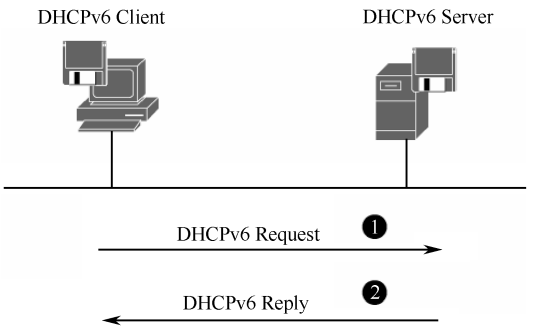

图 2-7　DHCPv6 基本过程

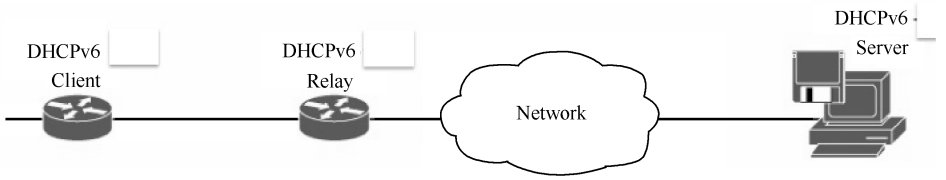

图 2-8　中继基本过程

思科采纳链路层地址加时间类型，并使用最小编号接口的 MAC 地址形成 DUID。DUID-LLT 的格式如图 2-9 所示。

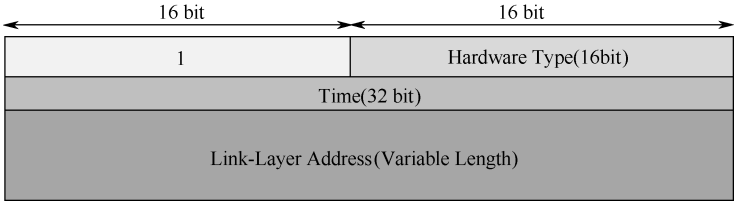

图 2-9　DUID-LLT 格式

- 最高 2 字节为类型字段：值为 1；
- 随后 2 字节是链路层地址类型字段：以太地址使用类型 1；
- 4 字节时间字段：以秒为单位表示生成 DUID 时的时间，计算公式是从 2000 年 1 月 1 日 0 时（UTC）起，以 2^{32} 取模；
- 最后 4 字节是链路层地址。

分配给客户端的一个或多个地址称为身份关联（Identity Association，IA），每个 IA 有一个 ID 号（IAID）。一个客户端可能有多个 IA，比如每个接口一个 IA。

服务器使用 DUID 确定为客户端分配哪些参数，也用来确定和客户端关联的 IA；当需要判断服务器身份时，客户端使用 DUID 来确定服务器。

客户端可以是主机或路由器的接口。客户端向服务器请求 IPv6 地址和其他配置参数，如 DNS、域名等。

客户端发送恳请（Solicit）消息查询可用的服务器，通过收集服务器的响应消息建立可用

服务器列表。服务器还可以在响应消息中设置优先级值，供客户端优先选择使用。

DHCP 服务器功能可以在启用了 IPv6 功能的路由器接口上启用。服务器为客户端提供地址前缀等其他配置参数信息，并在内存中维护着一个配置信息绑定表（Bingding），该表自动跟踪配置参数的变化，比如为客户端分配的地址前缀的变化等。这个自动的绑定表可以存储在 NVRAM 中，也可以存储在外部的 TFTP 服务器上。为维护方便，表中的记录以文本方式保存。

例如，每 120 秒向 TFTP 服务器写一次绑定信息：

```
ipv6 dhcp database tftp://10.0.0.1/dhcp-binding write-delay 120
```

每个配置池（Configuration Pool）关联一个绑定表，每条记录包含以下信息：

- DHCPv6 地址池名称；
- 接收到客户端请求消息的接口标识号；
- 客户端 IPv6 地址；
- 客户端 DUID；
- IAID；
- 前缀长度；
- 前缀的有效时间和超时时间；
- 其他参数。

服务器可以根据以上信息为客户端分配地址。

2.5.1　消息类型

DHCPv6 定义的消息类型有 13 种，其他未定义的留作未来使用。

- 恳请（Solicit）消息：类型=1，客户端用来查询服务器；
- 通告（Advertise）消息：类型=2，服务器响应恳请消息；
- 请求（Request）消息：类型=3，客户端向特定服务器请求地址等配置参数；
- 确认（Confirm）消息：类型=4，客户端向服务器确认其得到的地址对所在链路是否合适；
- 更新（Renew）消息：类型=5，客户端向服务器请示延长使用地址时间及更新其他参数；
- 再绑定（Rebind）消息：类型=6，当没收到服务器对更新消息的响应时，发送该消息；
- 应答（Reply）消息：类型=7，服务器回应客户端的恳请、请求、更新和再绑定消息，包含地址信息和其他配置参数，该消息也用来回应类型 4、8、9 和 11 消息。
- 释放（Release）消息：类型=8，客户端向服务器发送不再使用地址的消息。
- 拒绝（Decline）消息：类型=9，客户端用来通知服务器分配的地址已被占用。
- 再配置（Reconfigure）消息：类型=10，服务器用该消息通知客户端新增了配置参数或配置参数更新了，客户端需要更新其配置。
- 信息请求（Information-request）消息：类型=11，客户端向服务器请求除地址外的其他配置信息。
- 中继转发（Relay-forw）消息：类型=12，中继代理向服务器发送的信息。
- 中继应答（Relay-repl）消息：类型=13，服务器回应中继转发的消息，其中包含分配给客户端的地址等其他配置参数。

2.5.2　报文结构

客户端和服务器交互的所有 DHCP 消息都有相同的报文头，但选项部分的格式不固定。报文头结构如图 2-10 所示。

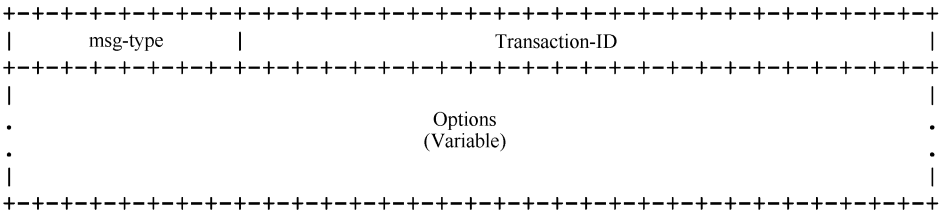

图 2-10　DHCP 报文头格式

- msg-type：消息类型，8 比特。
- Transaction-id：消息交互编号，24 比特。
- Option：选项，变长。

选项部分是用来携带具体的 DHCP 信息的字段，所有选项有着共同的格式，如图 2-11 所示。

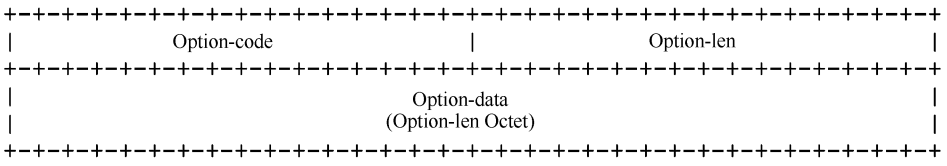

图 2-11　选项字段格式

- Option-code：选项代码；
- Option-len：选项长度；
- Option-data：选项携带的具体数据。

注：在 RFC3315 中记录的选项有 20 个，更多选项参考 RFC2132，本书不做详细讨论。

2.5.3　状态代码

状态代码用来描述 DHCP 交互的状态，包括以下几种状态。

- 成功（Success）：代码=0；
- 未知原因失败（UnspecFail）：代码=1；
- 无地址可用（NoAddrsAvail）：代码=2;
- 未绑定记录（NoBinding）：代码=3；
- 地址不合适（NotOnLink）：代码=4；对客户端所连结的链路来说，分配的地址是不合适的。
- 使用组播（UseMulticast）：代码=5；服务器向客户端发送的消息，强制客户端使用代

表 DHCP 服务器和中继的组播地址发送消息。

注：在记录的 20 个选项代码中，状态代码选项的编号为 13。

2.5.4　地址和端口号

DHCP 使用以下 2 个组播地址。

① FF02::1:2：该组播地址代表所有 DHCP 中继和服务器，客户端使用该地址与它们通信。

② FF05::1:3：该组播地址代表所有 DHCP 服务器，中继代理使用该地址查找并与服务器通信。

DHCP 使用 UDP 通信，客户端使用的端口号是 546，服务器和中继代理使用的端口号是 547。

2.5.5　启用 DHCP 服务

1. 在路由器上启用 DHCPv6 服务器

（1）定义地址池

```
Router(config)# ipv6 dhcp pool pool-name
Router(config-dhcpv6)# address prefix IPv6-prefix [lifetime {valid-lifetime preferred-lifetime | infinite}]
Router(config-dhcpv6)# link-address ipv6-prefix
Router(config-dhcpv6)# vendor-specific vendor-id
Router(config-dhcpv6-vs)#suboption number
```

- Ipv6-prefix：地址前缀；
- Link-address：客户端所在链路的地址，用以判断为客户端分配哪个地址池中的地址，参见中继代理部分；
- Vendor-specific：IANA 管理的厂商代码；
- Suboption：厂商定义的子选项编号，范围为 1～65535，可以是 IPv6 地址、ASCII 文本或十六进制字符串。

在地址池配置模式下，这些可选项的具体配置，包括 DNS、域名、Link-address、厂商自定义参数等。请读者参考命令手册。

例如：

```
Switch(config)# ipv6 dhcp pool testgroup
Switch(config-dhcpv6)# link-address 2001:1001::0/64
Switch(config-dhcpv6)# link-address 2001:1002::0/64
Switch(config-dhcpv6)# link-address 2001:2000::0/48
Switch(config-dhcpv6)# address prefix 2001:1003::0/64
Switch(config-dhcpv6)# vendor-specific 9
Switch(config-dhcpv6-vs)# suboption 1 address 1000:235D::1
Switch(config-dhcpv6-vs)# suboption 2 ascii "IP-Phone"
```

（2）在接口上启用 DHCPv6 服务器功能

```
Router(config-if)# ipv6 dhcp server [pool-name | automatic] [rapid-commit] [preference value] [allow-hint]
```

- Automatic：自动选择使用哪个地址池为客户分配地址。如果是直接（非中继转发）收到的 Silicit 报文，服务器通过比对报文入口地址和 Link-address 地址前缀或地址池前缀，判定使用哪个地址池，原则是最长匹配优先。
- 使用 Repaid-commit 参数，DHCPv6 客户端可以通过 2 个消息交换（Solicit、Reply）获得地址。默认情况下需要交换 4 个消息（Solicit、Advertise、Request、Reply）获得地址。
- Allow-hint：允许服务器分配 Solicit/Request 消息中暗示的有效地址。不使用该参数，服务器将忽略客户端的地址暗示。

例如：

```
Switch(config-if)# ipv6 dhcp server testgroup
```

检查 DHCPv6 运行状态的命令有：

（1）show ipv6 dhcp

```
Router# show ipv6 dhcp
This device's DHCPv6 unique identifier(DUID): 000300010002FCA5DC1C
```

（2）show ipv6 dhcp interface

```
Router1# show ipv6 dhcp interface
Ethernet2/1 is in server mode
  Using pool: svr-p1
  Preference value: 20
  Rapid-Commit is disabled
```

2. 启用接口客户端功能

```
Router(config-if)# ipv6 address dhcp    rapid-commit
```

如果需要获取厂商特定参数：

```
Router(config-if)# ipv6 dhcp client request vendor-specific
```

查看接口状态：

```
Router2# show ipv6 dhcp interface
Ethernet2/1 is in client mode
  State is OPEN (1)
  List of known servers:
    Address: FE80::202:FCFF:FEA1:7439, DUID 000300010002FCA17400
    Preference: 20
      IA PD: IA ID 0x00040001, T1 120, T2 192
        Prefix: 3FFE:C00:C18:1::/72
                preferred lifetime 240, valid lifetime 54321
                expires at Nov 08 2002 09:10 AM (54319 seconds)
```

```
        Prefix: 3FFE:C00:C18:2::/72
                preferred lifetime 300, valid lifetime 54333
                expires at Nov 08 2002 09:11 AM (54331 seconds)
        Prefix: 3FFE:C00:C18:3::/72
                preferred lifetime 280, valid lifetime 51111
                expires at Nov 08 2002 08:17 AM (51109 seconds)
      DNS server: 2001:0DB8:1001::1
      DNS server: 2001:0DB8:1001::2
      Domain name: example1.net
      Domain name: example2.net
      Domain name: example3.net
    Prefix name is cli-p1
    Rapid-Commit is enabled
```

如果让接口重新获取地址，需要重启接口客户端功能，例如，让 E0/0 接口重新获取地址：

```
Router# clear ipv6 dhcp client Ethernet 1/0
```

2.5.6　中继代理与服务器使用的消息格式

中继代理和服务器交互的消息有中继转发消息和中继应答消息两种，它们具有相同的报文格式，如图 2-12 所示。

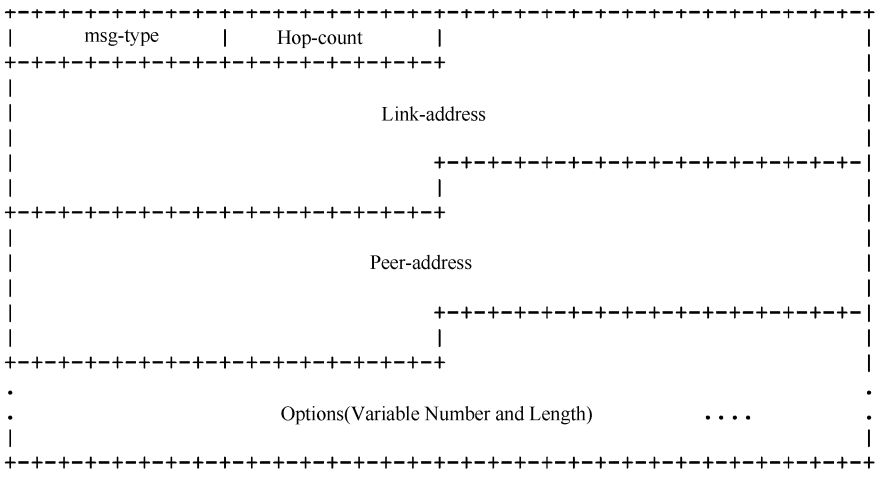

图 2-12　中继代理与服务器使用的消息格式

- msg-type：Relay-forw / Relay-reply；
- Hop-count：转发中继消息的中继代理数量；
- Link-address：接收到恳请消息的中继代理的接口地址，服务器用来判断客户端所在链路，选择地址池；
- Peer-address：中继代理转发中继消息的接口地址；
- Options：中继消息内容，格式如图 2-13 所示；
- Option_relay_msg：用选项代码 9 表示；

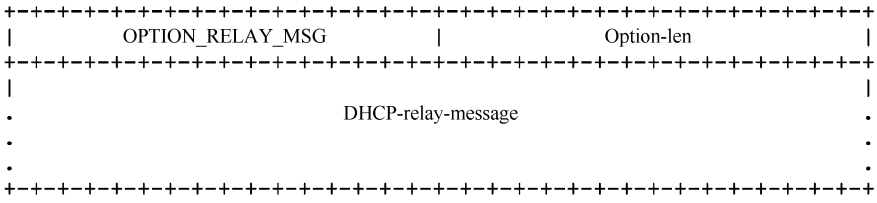

图 2-13　中继消息选项字段的格式

- Option-len：选项长度；
- DHCP-relay-message：在 Relay-forw 消息中是被中继的原始消息，在 Relay-reply 消息中是发送给 Peer-address 代表的中继代理或客户端的内容。

2.5.7　配置中继代理

启用路由器的中继代理功能，只需在连接客户端的接口上声明 DHCPv6 服务器的地址即可，例如：

```
Router(config)# interface ethernet 4/2
Router(config-if) ipv6 dhcp relay destination FE80::250:A2FF:FEBF:A056    Ethernet 4/3
```

通常，中继代理使用发送 Relay-forwd 消息的接口地址作为源地址。例如，上例的配置中强制使用 Ethernet 4/3 作为发送中继消息的接口，而不是由路由表决定使用哪个出口。如果希望中继代理使用更稳定的环回接口（如 loopback 0）作为源地址，可以使用如下命令配置：

```
ipv6 dhcp relay source-interface loopback 0
```

该命令既可以在全局模式下配置，也可以在接口模式下配置，接口下配置的命令优先。

2.5.8　无状态 DHCPv6

无状态的 DHCPv6 是指 DHCPv6 服务器为节点分配 DNS、域名等信息，但不动态维护节点的任何状态。这种无状态的 DHCPv6 功能由路由器通告 RA 控制，启用该功能的路由器可以响应客户端发送的无状态 DHCPv6 请求消息。

配置示例如下：

（1）服务器端开启无状态 DHCPv6

```
Router(config)# ipv6 dhcp pool stateless-dhcp-pool
Router(config-dhcp) dns-server 2001:0DB8:3000:3000::42
Router(config-dhcp)# domain-name domain1.com
Router(config-dhcp)# exit
Router(config)# interface serial 3
Router(config-if)# ipv6 dhcp server stateless-dhcp-pool
Router(config-if)# ipv6 nd other-config-flag
```

“ipv6 nd other-config-flag”的意思是，在该接口发送的 RA 消息中设置了这个标识，告诉客户端需要向 DHCPv6 服务器获取其他（非地址）信息。

（2）客户端侧的配置

```
Router(config-if)# ipv6 address autoconfig
```

在接口上启用该命令后，该接口的地址根据服务器的 RA 消息自动配置；如果 RA 消息中设置了“other-config-flag”标识，那么接口将通过 DHCPv6 服务器获取其他配置信息。

2.6　DHCPv6 前缀代表

DHCPv6 前缀代表简称 DHCPv6-PD，是对 DHCPv6 的扩展。DHCPv6 为客户端分配地址、DNS 等信息。DHCPv6-PD 可以为 DHCPv6-PD 客户端分配一个网络号，然后，这个 DHCPv6-PD 客户端可以把这个网络再分成多个网段给它的接口，以及这些接口连接的网段中的主机或节点使用。

比如，将企业客户的出口路由器配置成 DHCPv6-PD 客户端，将 ISP 的路由器配置成 DHCPv6-PD 服务器。客户的出口路由器将从 ISP 路由器那里获取一个网段地址，然后再进一步划分成多个子网号后，分配给内部主机使用。

2.6.1　部署方式

DHCPv6-PD 的部署方式有以下 3 种形式。

第一种：ISP 的边界路由器用作 DHCPv6-PD 服务器，DHCPv6-PD 客户端和服务器直连，如图 2-14 所示。

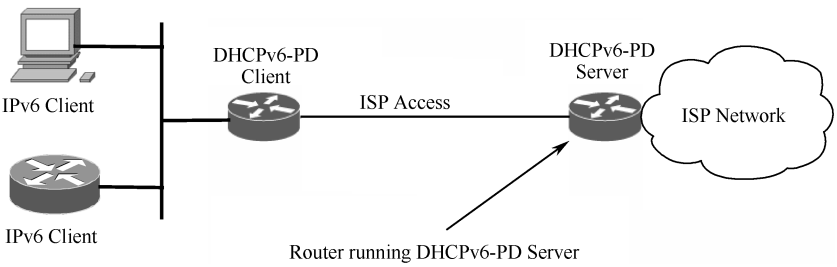

图 2-14　DHCPv6-PD 部署方式一

第二种：ISP 的边界路由器用作 DHCPv6-PD 中继，DHCPv6-PD 服务器在其他设备上，如图 2-15 所示。

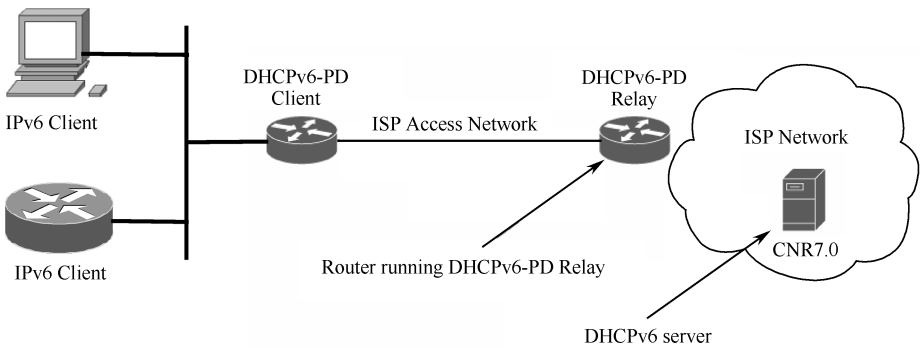

图 2-15　DHCPv6-PD 部署方式二

第三种：ISP 的边界路由器用作 DHCPv6-PD 中继，启用 remote-id 选项。这种部署模式和第二种相同，唯一不同的是如何向 IPv6 客户端分配 IPv6 地址。第二种方式是根据 DUID 为 IPv6 客户端分配地址，如果客户端由于某种原因连接到了其他 DHCPv6-PD 客户端上，地址分配和跟踪就比较复杂，管理系统也要更新，以适应客户端参数的变化。第三种方式是根据 Remote-ID 为 IPv6 客户端分配地址，这个 Remote-ID 可以指示 DHCPv6 服务器，最终客户端连接在哪个 DHCPv6-PD Client 上，基于选项中的一些信息分配地址，使得最终客户端可以独立于 DUID 而获取正确地址。

Remote-ID 是 DHCPv6 中继自动设置的一个值。

2.6.2　配置前缀代表

下面以图 2-16 所示的拓扑为例介绍如何配置前缀代表。

图中 DHCPv6-PD 服务器和 DHCPv6-PD 客户端直连。

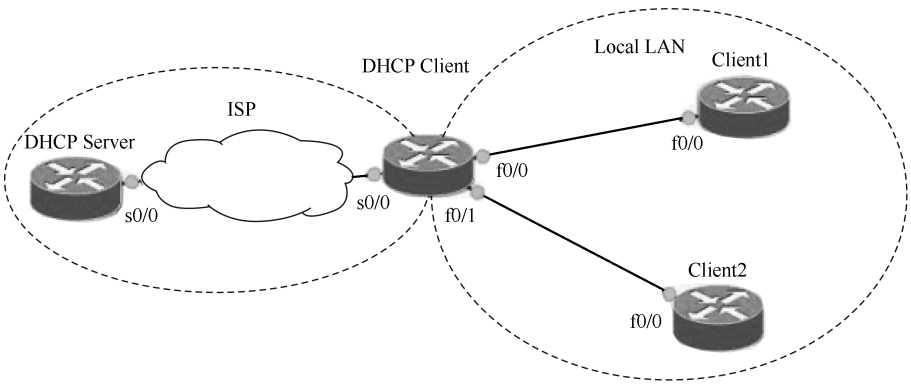

图 2-16　前缀代表拓扑

（1）DHCPv6-PD Server 的配置清单

```
hostname DHCP-PD-SERVER
!
ipv6 unicast-routing
ipv6 dhcp pool dhcpv6                    //定义服务 DHCP 的地址池名字
prefix-delegation pool dhcpv6-pool lifetime 1800 600          //调用名字为 dhcpv6-pool 的前缀代表地址池，并定义有效期
dns-server 2001:DB8:3000:3000::42
domain-name example.com
!
interface Serial0/0
 no ip address
 ipv6 address 2010:AB8:0:1::1/64
 ipv6 enable
 ipv6 dhcp server dhcpv6          //在接口上启用 DHCP 服务
```

```
 clock rate 2000000
!
ipv6 local pool dhcpv6-pool 2001:DB8:1200::/40 48          //定义一个前缀长度为 40 位的本地前缀代表地址池，并且定义分配给 DHCPv6-PD Client 的前缀是长度是 48 位
```

（2）DHCPv6-PD Client 的配置清单

```
hostname DHCPCLIENT
!
ipv6 unicast-routing
!
interface Serial0/0
no ip address
ipv6 address autoconfig default      //无状态自动配置 IPv6 地址，并生成一条指向 DHCPv6-PD 服务器的默认路由
!
ipv6 enable
ipv6 dhcp client pd prefix-from-provider     //在接口上启用 PD，并将 PD 命名为 prefix-from-provider
!
clock rate 2000000
!
interface FastEthernet0/0
no ip address
duplex auto
speed auto
ipv6 address prefix-from-provider ::1:0:0:0:1/64          //配置接口，从名字为 prefix-from-provider 的 PD 那里获取前缀，然后附加上（1:0:0:0:1），最后形成的接口地址是 2001:DB8:1200:1::1/64
ipv6 enable
!
interface FastEthernet0/1
no ip address
duplex auto
speed auto
ipv6 enable
ipv6 address prefix-from-provider ::1/64                  //接口地址为 2001:DB8:1200::1/64
```

（3）Client 1 的配置清单

```
hostname CLIENT1
!
ipv6 unicast-routing
!
interface FastEthernet0/0
```

```
no ip address
duplex auto
speed auto
ipv6 address autoconfig                    //启用无状态自动地址配置
ipv6 enable
```

（4）Client 2 的配置清单

```
hostname CLIENT2
!
ipv6 unicast−routing
!
interface FastEthernet0/0
no ip address
duplex auto
speed auto
ipv6 address autoconfig                    //启用无状态自动地址配置
ipv6 enable
```

2.6.3　验证 PD

（1）在 DHCPv6-PD Server 上查看地址池

```
DHCPv6-PD-SERVER# show ipv6 dhcp pool
DHCPv6 pool: dhcpv6
  Prefix pool: dhcpv6−pool
  preferred lifetime 600, valid lifetime 1800
  DNS server: 2001:DB8:3000:3000::42
  Domain name: example.com
  Active clients: 1
```

命令显示了地址池名字、前缀地址池名字、DNS、域名和当前的 DHCPv6-PD Client 的数量。

（2）在 DHCPv6-PD Server 上查看地址绑定信息

```
DHCPv6-PD-SERVER# show ipv6 dhcp binding
Client: FE80::C002:FFF:FEB4:0
  DUID: 00030001C2020FB40000
  Username : unassigned
  Interface : Serial0/0
  IA PD: IA ID 0x00060001, T1 300, T2 480
  Prefix: 2001:DB8:1200::/48
          preferred lifetime 600, valid lifetime 1800
      expires at Jan 26 2013 05:08 PM (1598 seconds)
```

（3）在 DHCPv6-PD Client 上查看启用 PD 的接口信息

```
DHCPCLIENT# show ipv6 dhcp interface
Serial0/0 is in client mode
  State is OPEN
  Renew will be sent in 00:04:37
  List of known servers:
  Reachable via address: FE80::C003:FFF:FEB4:0
  DUID: 00030001C2030FB40000
Preference: 0
Configuration parameters:
IA PD: IA ID 0x00060001, T1 300, T2 480
Prefix: 2001:DB8:1200::/48
        preferred lifetime 600, valid lifetime 1800
        expires at Mar 01 2002 10:59 AM (1777 seconds)
DNS server: 2001:DB8:3000:3000::42
Domain name: example.com
Information refresh time: 0
Prefix name: prefix−from−provider
Rapid−Commit: disabled
```

（4）查看接口 f0/0 获得的地址

```
DHCPCLIENT# show ipv6 interface f0/0
FastEthernet0/0 is up, line protocol is up
  IPv6 is enabled, link−local address is FE80::C002:FFF:FEB4:0
    Global unicast address(es):
    2001:DB8:1200:1::1, subnet is
          2001:DB8:1200:1::/64 [CAL/PRE]
      valid lifetime 1535 preferred lifetime 335
```

（5）查看接口 f0/1 获得的地址

```
DHCPCLIENT# show ipv6 interface f0/1
FastEthernet0/1 is up, line protocol is up
  IPv6 is enabled, link−local address is FE80::C002:FFF:FEB4:1
    Global unicast address(es):
    2001:DB8:1200::1, subnet is
          2001:DB8:1200::/64 [CAL/PRE]
```

（6）查看通过 PD 获得的前缀

```
DHCPCLIENT# show ipv6 general-prefix
IPv6 Prefix prefix−from−provider, acquired via DHCP PD
```

```
    2001:DB8:1200::/48 Valid lifetime 1656, preferred lifetime 456
     FastEthernet0/1 (Address command)
     FastEthernet0/0 (Address command)
```

（7）查看 Client 1 获得的地址

```
FastEthernet0/0 is up, line protocol is up
  IPv6 is enabled, link−local address
          is FE80::C000:FFF:FEB4:0
  Global unicast address(es):
  2001:DB8:1200:1:C000:FFF:FEB4:0, subnet is 2001:DB8:1200:1::/64 [EUI/CAL/PRE]
  valid lifetime 1709 preferred lifetime 509
```

（8）查看 Client 2 获得的地址

```
FastEthernet0/0 is up, line protocol is up
  IPv6 is enabled, link−local address
          is FE80::C001:FFF:FEB4:0
  Global unicast address(es):
2001:DB8:1200:0:C001:FFF:FEB4:0, subnet is 2001:DB8:1200::/64 [EUI/CAL/PRE]
valid lifetime 1770 preferred lifetime 570
```

2.7　IPv6 的移动特性

越来越多的用户并不总是在自己的办公桌上工作，他们的工作地点需要随时变动，甚至在旅途中办公。IPv6 提供对移动用户的支持是非常重要的，因为一台移动主机经常接入不同的网络，但不希望每次改变接入网络时更改自己在原网络使用的 IPv6 地址。

2.7.1　移动 IPv6 节点的构成部件

移动 IPv6 节点技术包含以下主要部件，如图 2-17 所示。

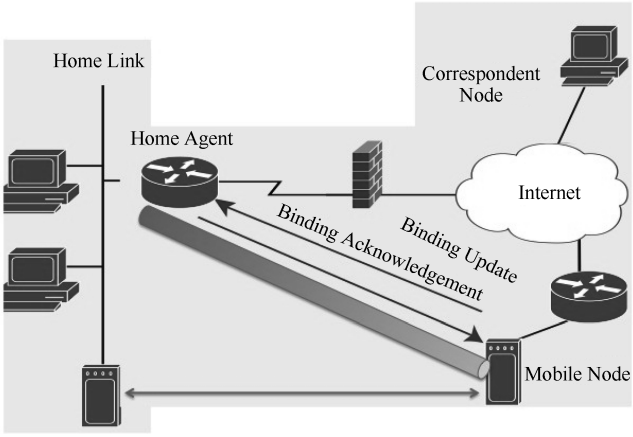

图 2-17　移动 IPv6 的部件

① Home Address——基地地址。移动节点在其原始网络上使用的地址。

② Mobile Node——移动节点。无论连接到任何网络上都使用自己的基地地址维护网络连接的主机。

③ Correspondent Node——通信者节点。与移动节点进行通信的目标主机。

④ Care of address（CoA）——转交地址。移动节点离开原始网络，当连接到其他网络时租借的地址。目标地址是移动节点基地地址的数据，需要由该地址转交给移动节点，即需要发送到该地址上。

⑤ Home Agent——基地代理。基地代理维护着移动节点基地地址和转交地址的映射关系。

2.7.2　移动 IPv6 节点的工作过程

当移动节点接入基地网络时获得一个基地地址，当移动节点漫游到其他网络时又会获得一个转交地址，因此，IPv6 的移动性管理就主要集中在基地地址和转交地址的映射关系管理上。管理基地地址和转交地址映射关系的工作由基地代理完成。

当移动节点漫游到其他网络时，它将通过路由器宣告信息获得一个转交地址。移动节点在每次改变其链路层的接入点时，即从一个 IPv6 子网移动到另一个 IPv6 子网时，都必须得到一个新的转交地址。以前获得的转交地址还可以保留，使移动节点仍然可以收到目标地址是以前转交地址的数据包。这一特性对无线移动节点非常有用，移动节点可以把自己配置到接收信号最强的子网，同时还可以接收原来为它提供服务的子网发出的信息。

基地地址和转交地址之间的映射关系称为绑定（Binding）。移动节点每次获得一个新的转交地址时，必须通过绑定更新信息（Binding Update，BU）将此地址发送给其基地代理。移动节点也能把自己当前的 BU 信息提供给对应节点，这个功能称为通信者注册（Correspondent Registration）。为完成该过程，通信者节点将进行反向路由可达性检测。

移动节点维护着 BU 列表，记录着每一个该节点发送的 BU 信息——那些发送给通行者节点的更新信息和那些发送给基地代理的更新信息。BU 列表中的条目都有超时时间的定义，过期的条目将被清除。

基地代理也维护着绑定更新（Binding Update，BU）信息，绑定更新信息由移动节点发送，用来通知移动节点当前位置的信息。基地代理为每个移动节点的每一个 IPv6 地址维护一个独立的绑定缓存（Binding Cache）。BU 列表包含所有移动节点发送的更新。

移动节点和通信者节点的通信模式有如下两种。

第一种模式称为双向隧道模式。这种模式不需要移动节点把自己的转交地址注册到通信者节点。如果移动节点不在基地网络中（已经漫游到其他网络），所有发送给移动节点的数据都将被基地代理拦截，然后，基地代理与移动节点建立一个隧道（如图 2-18 所示），把数据包转发给移动节点，移动节点的转交地址即为隧道的另一端点地址。

移动节点发送给通信者节点的数据包沿着反向隧道，即移动节点到基地代理的隧道，回到基地代理，然后由基地网络路由到通信者节点。在此模式下，基地代理使用代理 ND（Proxy ND）截获任何目标地址是移动节点基地地址的数据包，然后经由与移动节点的主地址建立的隧道发送给移动节点。

第二种模式称为路由优化模式。这种模式需要移动节点把自己的 BU 信息注册到通信者节点。这样，通信者节点发送给移动节点的数据包可以使用移动节点的转交地址作为目标地址。

当发送数据时，通信者节点查询缓存内的绑定信息，找出目标地址。

在封装 IPv6 报文时，路由扩展头的类型标记为 2，表示携带的是基地地址。类似地，移动节点使用转交地址作为源地址，在目标地址选项扩展头中使用选项类型 201，表示携带的是基地地址。

当路由器发送数据包时，首先从绑定缓存中查找目标地址，再从邻居发现协议建立的缓存中查找目标地址。绑定缓存中包含有为每一个移动节点基地地址建立的缓存条目。

当移动节点变动位置时，它将向绑定更新列表中列出的所有节点发送绑定更新信息。

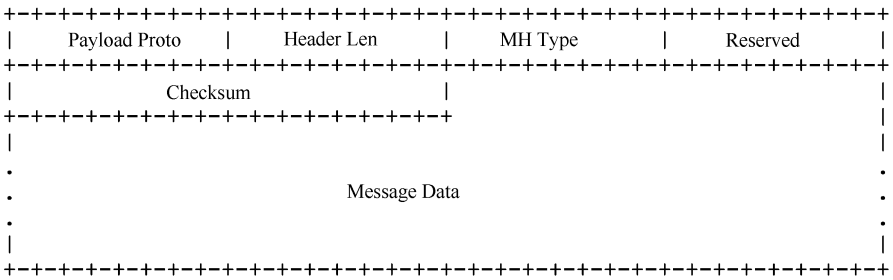

图 2-18　移动性扩展报文头格式

2.7.3　移动性扩展报文头

为支持移动 IPv6 特性，IPv6 协议定义了移动性扩展报文头（Mobility Header，MH），其代码为 135。移动节点、通信者节点和基地代理使用该扩展报文头建立和管理绑定信息。

移动性扩展报文头的格式如下，各字段含义如下所述。

- Payload Proto：8 比特长下一个报文头代码，代码=59。
- Header Len：8 比特报文头长度，必须是 8 字节的整数倍。
- MH Type：有关移动性的信息类型，8 比特长。
- Reserved：保留 8 比特。
- Checksum：报文头校验和，16 比特。
- Message Data：MH Type 字段定义的信息和可能有的选项，包括选项填充数据（如果存在），该字段为变长字段。

MH 类型如下所述。

- Binding Refresh Request：类型=0，请求移动节点更新其绑定信息的消息。
- Home Test Init：类型=1，移动节点发起，测试基地地址与通信者节点的可达性。
- Care-of Test Init：类型=2，移动节点发起，测试转交地址与通信者节点的可达性。
- Home Test：类型=3，通信者节点响应类型 1 的消息。
- Care-of Test：类型= 4，通信者节点响应类型 2 的消息。
- Binding Update：类型=5，移动节点向其他节点通知其新的转交地址的更新消息。
- Binding Acknowledgement：类型=6，响应 BU 的消息。
- Binding Error：类型=7，通信者节点发送的有关移动性的错误消息。

关于移动性报文的详细内容请参考 RFC 3775。

2.8　本 章 小 结

在 IPv6 环境里，已经没有了地址解析协议，取而代之的是 ICMPv6。除此之外，ICMPv6 协议还用来发现邻居、重定向等。

NDP 协议是用来发现对方链路层地址的协议，除此之外，其功能还包括前缀发现、自动地址配置、可达性检测等。NDP 协议是一个很重要的协议，它利用 ICMPv6 协议完成上述功能。本章详细介绍了 NDP 工作使用的消息类型，包括路由器宣告、路由器请求、邻居宣告、邻居请求和路由重定向。

DHCPv6 提供有状态的地址和其他信息分配。工作过程与 IPv4 的 DHCP 类似。本章详细介绍了 DHCPv6 报文结构、使用的 13 种消息类型和 6 种状态代码。另外，还介绍了 DHCPv6 的配置命令。

移动 IPv6 具有全新的特性，为移动节点提供了方便的通信功能。构成移动 IPv6 的元素包括移动节点、基地代理、通信者节点、基地地址和转交地址。这些元素间的映射关系由绑定表维护。移动 IPv6 工作的模式有双向隧道模式和路由优化模式两种。

第 3 章　支持 IPv6 的 RIP 协议

为支持 IPv6 而开发的 RIP 路由协议称为 RIPng，即下一代 RIP 的意思。RIPng 功能和 IPv4 协议中的 RIPv2 相同，也可以看作 RIPv2 的增强版——增强了支持 IPv6 地址和前缀等特性。RIPng 更新路由信息使用的地址也是组播地址，为其分配的是 FF02::9，代表所有运行 RIPng 协议的路由器。

和 RIPv2 协议相比，它们有许多的相同特性，例如：

- 最大 15 跳的限制；
- 使用 UDP 传送路由信息；
- 都是距离矢量型路由协议；
- 30 秒的定期更新周期；
- 定期更新和触发更新并用；
- 遵守 Split Horizon 规则；
- 遵守 Poison Reverse 规则；
- 同样具有环路和计算到无穷的风险。

运行 IOS 的路由器上可以运行多个 RIPng 进程。路由器的接口可以运行在不同的进程，也就是说，可以在不同的进程下与不同的邻居交换路由信息。

每一个 RIPng 进程维护一个本地的路由表，称为路由信息库（Routing Information Database，RIB），RIB 中包含从邻居学习到的度量最小的路由条目。RIB 中也存储那些正向邻居通告的过期的路由条目。

RIPng 把那些本地 RIB 中未过期的路有条目安装在主 RIB（Master RIB）中。

3.1　RIPng 消息格式

RIPng 使用 UDP 发送和接受路由信息，源端口号和目标端口号都是 521。UDP 报文的数据字段部分携带的是路由信息，消息报文格式如图 3-1 所示。

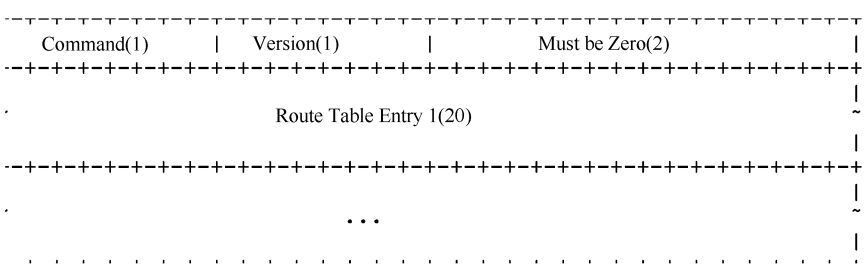

图 3-1　RIPng 消息格式

其中，

- Command：占 1 字节，表示消息类型、请求消息或响应消息。
- Version：占 1 字节，版本号，当前为版本 1。
- 紧随版本号字段的 2 个字段为 0。
- Route Table Entry：携带的具体路由条目，每个路由条目占 20 字节。根据 MTU 值可以计算出一个 UDP 报文可以携带多少条路由。

3.1.1　路由条目格式

每个路由条目包含前缀、前缀长度、度量值等，格式如图 3-2 所示。

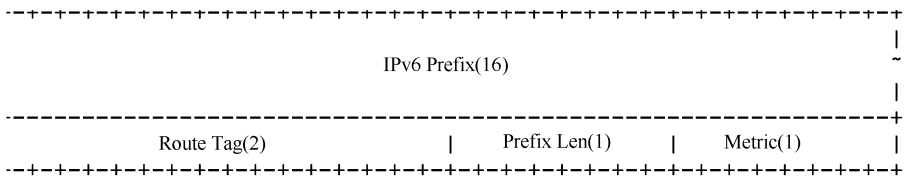

图 3-2　路由条目格式

其中，

- IPv6 Prefix：占 16 字节的路由前缀。
- Route Tag：占 2 字节的可能存在的路由标记。
- Prefix Len：占 1 字节的前缀长度。
- Metric：占 1 字节的度量值。

3.1.2　下一跳属性值

每一条路由都有下一跳属性，在 RIPv2 中，每条路由条目后面都有下一跳字段，用来携带下一跳地址。如果在 RIPng 中也如此实现的话，每条路由占用的字段几乎要多一倍（多占用 16 字节），这样导致传送路由信息的效率大大降低。因此，RIPng 并不这么做，它的做法是在所有路由条目最后携带一个下一跳，表明这些路由条目来自同一个地址，即来自相同地址的路由条目使用一个下一跳地址。

下一跳地址字段是在标准的路由条目字段上，通过把 Metric 字段设置为 0xFF 来表示。前缀字段携带的即为下一跳地址，Tag 字段和 Prefix Length 字段必须设置为 0。

3.1.3　请求消息

请求消息用来向邻居请求全部或部分路由信息。比如路由器刚启用 RIPng，为了快速获得路由信息，路由器使用组播地址向特定的 UDP 端口发送请求信息。为了表示是请求完整的路由信息而不是特定的路由信息，必须将路由前缀字段和前缀长度字段设置为 0，而将度量字段设置为无穷大（比如，16）。路由器收到这样的请求消息后，将把自己的全部路由响应给对方。

如果是请求特定的路由，路由器逐条检查消息中的路由条目，如果自己有该条目，则修改其度量值字段；如果没有则把度量值字段设置为无穷大，然后以响应消息的类型发回请求者。

3.1.4　响应消息

路由器在不同情况下都可以发送响应消息，比如对特定请求消息的回应，定期路由信息更新和由于路由变化而导致的触发更新。

路由器收到响应消息后，需要对其进行有效性验证，例如，是不是来自 RIPng 的既定端口，是不是来自有效的邻居，数据包的源地址是不是本链路地址，是否是自己发送的消息。如果这些项目不正确，则忽略此消息。

除此之外，路由器还需要检查 TTL 是否是 255，这可以验证数据是不是来自邻居。因为，如果是由中间人转送过来的，TTL 值应该被减小，不是 255。

报文的完整性被检查后，还要检查路由条目项，比如 Metric 是否是无穷大，前缀地址是否有效，前缀长度是否有效。任何无效项目即导致忽略此路由条目。

这一切验证通过后，才进入加入路由表的过程。

3.2　配置 RIPng

配置 RIPng 需要做以下 4 方面的工作：

第一，全局启用转发 IPv6 单播特性。

第二，启用 RIPng 进程。一台路由器上可以启用多个进程，用不同名字区分。

第三，在接口模式下，把接口运行在 RIPng 进程中。

第四，其他优化配置。比如调整计时器、过滤路由、路由重分布等。

3.2.1　配置命令解释

（1）全局启用转发 IPv6 单播

```
Router(config)# ipv6 unicast-routing
```

（2）启用进程

```
Router(config)# ipv6 router rip process-name
```

用不同的名字代表不同的进程。

（3）把接口运行在进程中

```
Router(config-if)# ipv6 rip process-name enable
```

这样，可以把接口运行在不同的进程中。

（4）调整等值路径负载均衡

```
Router(config-rtr)# maximum-paths number-path
```

路由器支持的最大等值路径条数是 64，默认值是 4。

（5）产生默认路由

Router(config-if)# **ipv6 rip** *name* **default-information {only | originate} [metric** *metric-value*]

- Only：在该接口下只发送默认路由（::/0）。
- Originate：默认路由和其他路由一同发送。
- Metric：为默认路由设置的度量值。

（6）配置路由重分布

把其他路由信息重分布到 RIPng 进程中。

Router(config)# **ipv6 router rip** *process-name*

Router(config-rtr)# **redistribute** *protocol* [*process-id*] {**level-1 | level-1-2 | level-2**} [**metric** *metric-value*] [**metric-type** {**internal | external**}] [**route-map** *map-name*]

- Protocol：变量不仅指动态路由协议，还可以是静态（Static）路由和直连（Connected）路由。
- Rout-map 的语法和含义与 IPv4 下的相同，只不过是用 IPv6 地址替换 IPv4 地址。

（7）使用分布列表过滤路由

IPv6 分布列表（Distribute-list）可以让路由器过滤接收到的或发送出去的路由，分布列表既可以应用在路由进程模式下，也可以应用在特定的接口上。

分布列表调用前缀列表，前缀列表中定义的路由前缀是用来过滤路由的条件。前缀列表中允许的路由前缀，可以被发送或接收。

前缀列表的语法如下：

Router(config)# **ipv6 prefix-list** *prefix-list-name* [**seq** *seq-number*] {**deny** *ipv6-prefix/prefix-length* | **description** *text*} [**ge** *ge-value*] [**le** *le-value*]

Router(config)# **ipv6 prefix-list** *prefix-list-name* [**seq** *seq-number*] {**permit** *ipv6-prefix/prefix-length* | **description** *text*} [**ge** *ge-value*] [**le** *le-value*]

- Le：前缀长度小于等于该值；
- Ge：前缀长度大于等于该值；
- 如果不使用 ge 或 le，则精确匹配 prefix-length 值；

关于前缀长度的约定：*prefix-length* < ge≤le≤128。

（8）分布列表应用在路由进程下

如果应用在路由进程模式下，且没有声明接口，则适用于所有运行在该进程中的接口。如果声明接口，则仅对该接口通告或接收的路由信息有效。

Router(config-rtr-rip)#**distribute-list prefix-list** *prefix-list-name* {**in** | **out**} *[interface-type interface-number]*

3.2.2　配置示例

该示例显示的是在路由器上启用了一个名字叫 process1 的 IPv6 进程，并调整等值路径的数量为 1。

把 BGP65001 的路由信息重分布到 RIPng 中，并使用路由图进行过滤。路由图调用的前缀列表对重分布的路由进行过滤。同时，为重分布进入 RIPng 的路由都设置标记（Tag）为 4。

在进程下，还使用分布列表对从接口 E0/0 进入的路由信息过滤：禁止该接口接收默认路由，允许接收其他所有路由。因为在该接口上使用 default-information 命令向 RIPng 进程中注入了默认路由，这样避免路由回馈。

配置清单如下：

```
ipv6 router rip process1
maximum-paths 1
redistribute bgp 65001 route-map bgp-to-rip
distribute-list prefix-list eth0/0-in-flt in Ethernet0/0
!
interface Ethernet0/0
ipv6 address 2001:0db8::/64 eui-64
ipv6 rip process1 enable
ipv6 rip process1 default-information originate
!
ipv6 prefix-list bgp-to-rip-flt seq 10 deny 2001:0db8:3::/16 le 128
ipv6 prefix-list bgp-to-rip-flt seq 20 permit 2001:0db8:1::/8 le 128
!
ipv6 prefix-list eth0/0-in-flt seq 10 deny ::/0
ipv6 prefix-list eth0/0-in-flt seq 15 permit ::/0 le 128
!
route-map bgp-to-rip permit 10
match ipv6 address prefix-list bgp-to-rip-flt
set tag 4
```

3.2.3　查看运行状态的命令

（1）查看正在运行的 RIPng 进程

```
Router> show ipv6 rip
RIP process "process1", port 521, multicast-group FF02::9, pid 62
      Administrative distance is 120. Maximum paths is 1
      Updates every 30 seconds, expire after 180
      Holddown lasts 0 seconds, garbage collect after 120
            Split horizon is on; poison reverse is off
      Default routes are generated
      Periodic updates 223, trigger updates 1
        Interfaces:
    Ethernet0/0
        Redistribution:
    Redistributing protocol bgp 65001 route-map bgp-to-rip
```

输出显示，当前运行的进程是“process1”，使用组播地址 FF02::9、端口号 521 通告路由信息。产生了默认路由，重分布进入了 BGP 路由，并应用了路由图 bgp-to-rip 对重分布进行了过滤。

（2）查看路由信息库

```
Router> show ipv6 rip process1 database
       RIP process "process1", local RIB
 2001:0db8::/64, metric 2
              Ethernet0/0/FE80::A8BB:CCFF:FE00:B00, expires in 13 secs
 2001:0db8::/16, metric 2 tag 4, installed
              Ethernet0/0/FE80::A8BB:CCFF:FE00:B00, expires in 13 secs
 2001:0db8:1::/16, metric 2 tag 4, installed
              Ethernet0/0/FE80::A8BB:CCFF:FE00:B00, expires in 13 secs
 2001:0db8:2::/16, metric 2 tag 4, installed
              Ethernet0/0/FE80::A8BB:CCFF:FE00:B00, expires in 13 secs
 ::/0, metric 2, installed
              Ethernet0/0/FE80::A8BB:CCFF:FE00:B00, expires in 13 secs
```

- Ethernet0/0/FE80::A8BB:CCFF:FE00:B00：表示学习到路由的接口和邻居地址。
- Installed：表示该路由被安装到 IPv6 路由表中了。

（3）查看从接口和邻居学习到的路由总数

```
Router> show ipv6 rip process1 next-hops
RIP process "process1", Next Hops
          FE80::A8BB:CCFF:FE00:A00/Ethernet0/0 [4 paths]
```

使用 next-hops 参数可以查看从某接口和邻居处学习到的路由总数，此例显示从 E0/0 接口的邻居那里学习到 4 条路由。

FE80::A8BB:CCFF:FE00:A00：邻居的地址。

注意：如果邻居使用接口地址而非本链路地址通告路由，此处不显示地址。

（4）查看路由表

```
Router> show ipv6 route rip
        IPv6 Routing Table - 17 entries
        Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
               U - Per-user Static route
               I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea
               O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
        R      2001:0db8:1::/32 [120/2]
               via FE80::A8BB:CCFF:FE00:A00, Ethernet0/0
        R      2001:0db8:2::/32 [120/2]
               via FE80::A8BB:CCFF:FE00:A00, Ethernet0/0
        R      2001:0db8:3::/32 [120/2]
```

```
                via FE80::A8BB:CCFF:FE00:A00, Ethernet0/0
```

（5）查看路由信息交换过程

```
Router# debug ipv6 rip
RIPng: Sending multicast update on Ethernet0/0 for process1
        src=FE80::A8BB:CCFF:FE00:B00
        dst=FF02::9 (Ethernet0/0)
        sport=521, dport=521, length=112
        command=2, version=1, mbz=0, #rte=5
        tag=0, metric=1, prefix=2001:0db8::/64
        tag=4, metric=1, prefix=2001:0db8:1::/16
        tag=4, metric=1, prefix=2001:0db8:2;:/16
        tag=4, metric=1, prefix=2001:0db8:3::/16
        tag=0, metric=1, prefix=::/0
RIPng: Next RIB walk in 10032
RIPng: response received from FE80::A8BB:CCFF:FE00:A00 on Ethernet0/0 for process1
        src=FE80::A8BB:CCFF:FE00:A00 (Ethernet0/0)
        dst=FF02::9
        sport=521, dport=521, length=92
        command=2, version=1, mbz=0, #rte=4
        tag=0, metric=1, prefix=2001:0db8::/64
        tag=0, metric=1, prefix=2001:0db8:1::/32
        tag=0, metric=1, prefix=2001:0db8:2::/32
        tag=0, metric=1, prefix=2001:0db8:3::/32
```

第一部分显示的是从接口 E0/0 发送的路由进程“process1”的路由信息，如下所述。

- 源地址：src=FE80::A8BB:CCFF:FE00:B00
- 目标地址：dst=FF02::9
- 源端口号：sport=521
- 目标端口号：dport=521
- 数据长度：length=112

后面是具体通告的路由条目和路由的度量值及标记（Tag）。

第二部分显示的是接收到的路由信息。

3.3　本 章 小 结

本章介绍了下一代 RIP 路由协议——RIPng 的工作原理，并和 IPv4 的 RIP 路由协议进行了对比。在介绍了基本工作原理和特点后，详细介绍了 RIPng 使用的消息格式、路有条目格式、请求消息和响应消息。

学习了原理之后，本章 3.2 节介绍了配置 RIPng 的命令语法，并解释了命令的作用。最后举例介绍了 RIPng 的配置过程，并介绍了检查 RIPng 运行的命令及其输出的信息含义。

第 4 章　支持 IPv6 的 EIGRP 协议

EIGRP 路由协议是思科在 IGRP 协议的基础上开发的性能更优的路由协议，该协议采纳斯坦福国际研究院（SRI International）发明的收敛技术（Convergence Technology），并使用了一种称为扩散更新算法（Diffusing Update Algorithm，DUAL）的路由计算方法。这种算法在保证路由无环的前提下，尽可能减少参与运算的路由器数量：仅计算拓扑变化影响的路由器，没有受到拓扑变化影响的路由器不参与路由再计算。

在 IPv6 环境下，EIGRP 的运行过程和 IPv4 下的过程一样。EIGRP 使用新的 TLV 携带 IPv6 可达性信息（路由信息）。

支持 IPv6 的 EIGRP 也有不同与 IPv4 环境下的特征：

- EIGRP IPv6 进程直接运行在接口上，并没有类似 IPv4 下的 network 命令。只要接口下启用了 EIGRP 协议，不需要在路由器模式下配置任何命令 EIGRP 就可以运行。
- EIGRP 需要路由器 ID 号。
- 接口在 Passive 模式下不需要配置 IPv6 进程。
- EIGRPv6 进程可以“Shut Down”。

4.1　配置 EIGRPv6 的命令

（1）启用 IPv6 单播路由功能

```
Router(config)# ipv6 unicast-routing
```

（2）在全局启用 EIGRP 进程

```
Router(config)# ipv6 router eigrp as-number
```

（3）配置路由器 ID

```
Router(config-rtr)# router-id ip-addres
```

- Ip-address：用 IPv4 地址作为 ID 号。

（4）在接口上启用 EIGRP 进程

```
Router(config-if)# ipv6 enable
Router(config-if)# ipv6 eigrp as-number
```

上述命令在配置清单中显示如下：

```
ipv6 unicast-routing
interface gigabitethernet0/0
```

```
 ipv6 enable
 ipv6 eigrp 1
 no shut
!
ipv6 router eigrp 1
 router-id 10.1.1.1
 shutdown
```

4.1.1　配置示例

下面以图 4-1 为例配置 EIGRPv6。

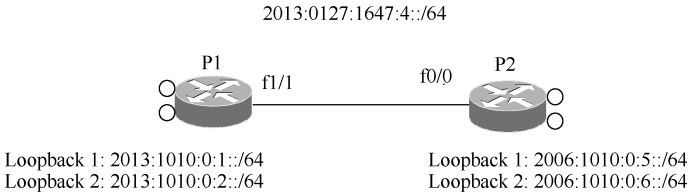

图 4-1　配置 EIGRPv6

（1）配置路由器 P1

```
P1#config t
P1(config)#ipv6 unicast-routing
P1(config)#ipv6 router eigrp 88
P1(config-rtr)#router-id 1.1.1.1
P1(config-rtr)#no shutdown
P1(config-rtr)#exit
P1(config)#interface f1/1
P1(config-if)#ipv6 address 2013:0127:1647:4::/64 eui-64
P1(config-if)#ipv6 enable
P1(config-if)# ipv6 eigrp 88
P1(config-if)#exit
P1(config)#interface loopback 1
P1(config-if)#ipv6 address 2013:1010:0:1::/64 eui-64
P1(config-if)#ipv6 enable
P1(config-if)#ipv6 eigrp 88
P1(config-if)#exit
P1(config-if)#interface loopback 2
P1(config-if)#ipv6 add 2013:1010:0:2::/64 eui-64
P1(config-if)#ipv6 enable
P1(config-if)#ipv6 eigrp 88
P1(config-if)#exit
P1(config)#
```

（2）配置路由器 P2

```
P2#config t
P2(config)#ipv6 unicast-routing
P2(config)#ipv6 router eigrp 88
P2(config-rtr)#router-id 2.2.2.2
P1(config-rtr)#no shutdown
P2(config-rtr)#exit
P2(config)#interface f0/0
P2(config-if)#ipv6 address 2013:0127:1647:4::/64 eui-64
P1(config-if)#ipv6 enable
P2(config-if)# ipv6 eigrp 88
P1(config-if)#exit
P2(config)#interface loopback 1
P2(config-if)#ipv6 address 2006:1010:0:5::/64 eui-64
P1(config-if)#ipv6 enable
P2(config-if)#ipv6 eigrp 88
P1(config-if)#exit
P2(config-if)#interface loopback 2
P2(config-if)#ipv6 add 2006:1010:0:6::/64 eui-64
P1(config-if)#ipv6 enable
P2(config-if)#ipv6 eigrp 88
P2(config-if)#exit
P2(config)#
```

4.1.2　验证

（1）查看运行的路由协议

```
P2#sh ipv6 protocols
IPv6 Routing Protocol is "connected"
IPv6 Routing Protocol is "static"
IPv6 Routing Protocol is "eigrp 88"
  EIGRP metric weight K1=1, K2=0, K3=1, K4=0, K5=0
  EIGRP maximum hopcount 100
  EIGRP maximum metric variance 1
  Interfaces:
    FastEthernet0/0
    Loopback1
    Loopback2
  Redistribution:
    None
```

```
    Maximum path: 16
    Distance: internal 90 external 170
```

此命令显示路由器上正在运行的路由协议 AS 号、哪些接口运行在进程里等。

（2）查看邻居

```
P2#sh ipv6 eigrp neighbors
IPv6-EIGRP neighbors for process 88
H   Address                 Interface       Hold Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
0   Link-local address:     Fa0/0           13 00:12:55   30     200  0   21
    FE80::C803:49FF:FE21:1D

P1#sh ipv6 eigrp nei
IPv6-EIGRP neighbors for process 88
H   Address                 Interface       Hold Uptime   SRTT   RTO  Q  Seq
                                            (sec)         (ms)       Cnt Num
0   Link-local address:     Fa1/1           11 00:15:35   31     200  0   18
    FE80::C806:49FF:FE46:8
```

输出显示，邻居关系是依靠本链路地址建立的。

（3）查看拓扑库

```
P1#sh ipv6 eigrp topology
IPv6-EIGRP Topology Table for AS(88)/ID(1.1.1.1)

Codes: P - Passive, A - Active, U - Update, Q - Query, R - Reply,
       r - reply Status, s - sia Status

P 2013:1010:0:2::/64, 1 successors, FD is 128256
        via Connected, Loopback2
P 2013:1010:0:1::/64, 1 successors, FD is 128256
        via Connected, Loopback1
P 2006:1010:0:6::/64, 1 successors, FD is 156160
        via FE80::C806:49FF:FE46:8 (156160/128256), FastEthernet1/1
P 2006:1010:0:5::/64, 1 successors, FD is 156160
        via FE80::C806:49FF:FE46:8 (156160/128256), FastEthernet1/1
P 2013:127:1647:4::/64, 1 successors, FD is 28160
        via Connected, FastEthernet1/1
```

输出显示，下一跳地址都使用本链路地址。

（4）查看 EIGRPv6 相关数据

```
P2#sh ipv6 eigrp traffic
```

```
IPv6-EIGRP Traffic Statistics for AS 88
  Hellos sent/received: 2758/2188
  Updates sent/received: 24/20
  Queries sent/received: 4/0
  Replies sent/received: 0/4
  Acks sent/received: 16/18
  Input queue high water mark 2, 0 drops
  SIA-Queries sent/received: 0/0
  SIA-Replies sent/received: 0/0
  Hello Process ID: 227
  PDM Process ID: 226
```

输出显示发送和接收到的 Hello 和 Update 等数据信息。

（5）查看路由表

```
P2#sh ipv6 route
IPv6 Routing Table - 9 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       U - Per-user Static route, M - MIPv6
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
       D - EIGRP, EX - EIGRP external
C    2006:1010:0:5::/64 [0/0]
      via ::, Loopback1
L    2006:1010:0:5:C806:49FF:FE46:8/128 [0/0]
      via ::, Loopback1
C    2006:1010:0:6::/64 [0/0]
      via ::, Loopback2
L    2006:1010:0:6:C806:49FF:FE46:8/128 [0/0]
      via ::, Loopback2
C    2013:127:1647:4::/64 [0/0]
      via ::, FastEthernet0/0
L    2013:127:1647:4:C806:49FF:FE46:8/128 [0/0]
      via ::, FastEthernet0/0
D    2013:1010:0:1::/64 [90/156160]
      via FE80::C803:49FF:FE21:1D, FastEthernet0/0
D    2013:1010:0:2::/64 [90/156160]
      via FE80::C803:49FF:FE21:1D, FastEthernet0/0
L    FF00::/8 [0/0]
      via ::, Null0
```

```
P1#sh ipv6 route
IPv6 Routing Table - 9 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       U - Per-user Static route, M - MIPv6
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
       D - EIGRP, EX - EIGRP external
D      2006:1010:0:5::/64 [90/156160]
       via FE80::C806:49FF:FE46:8, FastEthernet1/1
D      2006:1010:0:6::/64 [90/156160]
       via FE80::C806:49FF:FE46:8, FastEthernet1/1
C      2013:127:1647:4::/64 [0/0]
       via ::, FastEthernet1/1
L      2013:127:1647:4:C803:49FF:FE21:1D/128 [0/0]
       via ::, FastEthernet1/1
C      2013:1010:0:1::/64 [0/0]
       via ::, Loopback1
L      2013:1010:0:1:C803:49FF:FE21:6/128 [0/0]
       via ::, Loopback1
C      2013:1010:0:2::/64 [0/0]
       via ::, Loopback2
L      2013:1010:0:2:C803:49FF:FE21:6/128 [0/0]
       via ::, Loopback2
L      FF00::/8 [0/0]
       via ::, Null0
```

注：FF00::/8 的路由是自动产生的。

（6）使用 debug 命令查看路由信息更新过程

首先在 P1 上打开 debug：debug ipv6 eigrp，然后把 P2 的 f0/0 接口关闭再开启，在 P2 上可以看到如下的信息：

```
P1#
*Jan 27 18:38:27.287: %DUAL-5-NBRCHANGE: IPv6-EIGRP(0) 88: Neighbor FE80::C806:49FF:FE46:8 (FastEthernet1/1) is up: new adjacency   //邻居关系建立
*Jan 27 18:38:27.303: IPv6-EIGRP(0:88): 2013:127:1647:4::/64 - do advertise out FastEthernet1/1 //通告路由
*Jan 27 18:38:27.303: IPv6-EIGRP(0:88): Int 2013:127:1647:4::/64 metric 28160 - 25600 2560
*Jan 27 18:38:27.303: IPv6-EIGRP(0:88): 2013:1010:0:1::/64 - do advertise out FastEthernet1/1
*Jan 27 18:38:27.303: IPv6-EIGRP(0:88): Int 2013:1010:0:1::/64 metric 128256 - 256 128000
*Jan 27 18:38:27.303: IPv6-EIGRP(0:88): 2013:1010:0:2::/64 - do advertise out FastEthernet1/1
*Jan 27 18:38:27.303: IPv6-EIGRP(0:88): Int 2013:1010:0:2::/64 metric 128256 - 256 128000
```

```
*Jan 27 18:38:29.327: IPv6-EIGRP(0:88): Processing incoming UPDATE packet    //处理接受的邻居通告的路由
*Jan 27 18:38:29.327: IPv6-EIGRP(0:88): Int 2006:1010:0:5::/64 M 156160 - 25600 130560 SM 128256 - 256 128000
*Jan 27 18:38:29.327: IPv6-EIGRP(0:88): 2006:1010:0:5::/64 (90/156160) added to RIB
*Jan 27 18:38:29.327: IPv6-EIGRP(0:88): Int 2006:1010:0:6::/64 M 156160 - 25600 130560 SM 128256 - 256 128000
*Jan 27 18:38:29.327: IPv6-EIGRP(0:88): 2006:1010:0:6::/64 (90/156160) added to RIB
*Jan 27 18:38:29.327: IPv6-EIGRP(0:88): Int 2013:127:1647:4::/64 M 30720 - 25600 5120 SM 28160 - 25600 2560
*Jan 27 18:38:29.327: IPv6-EIGRP(0:88): 2013:127:1647:4::/64 routing table not updated
```

输出显示，有 2 条路由加入路由信息库（RIB），有一条路由没有在路由表里更新（Routing Table Not Updated）。

4.2　调整相关参数

（1）调整 EIGRP 使用的接口带宽值

默认地，EIGRP 报文最多占用接口声称带宽值的 50%，即 Bandwidth 声称带宽值的 50%。如果需要修改该值，可以使用如下命令：

```
Router(config-if)# ipv6 bandwidth-percent eigrp as-number percent
```

- *Percent*：接口带宽百分比。

（2）调整 Hello 间隔时间

路由器定期向直连的邻居发送 Hello 包，以维持邻居关系和判断邻居是否还在线工作。可以调整发送 Hello 包的时间周期。

```
Router(config-if)#ipv6 hello-interval eigrp as-number seconds
```

- *as-number*：AS 号；
- *seconds*：Hello 包时间间隔，以秒为单位。

（3）调整保持时间（Hold Time）

保持时间是路由器用来判定邻居是否存在的最大等待时间，即在这个等待时间段内，即便收不到邻居的 Hello 包，也认为邻居依然存在，超过这个 Hold 时间没收到邻居的 Hello，才认为邻居失效。默认的保持时间是 Hello 间隔时间的 3 倍。在低速 NBMA 链路上，默认保持时间是 180 秒。在大型网络或网络经历拥塞时，默认的保持时间可能不足以保证路由器收到所有邻居的 Hello 包，在这种情况下，需要调整保持时间。如果 Hello 间隔时间调整了，也应该调整保持时间。

```
Router(config-if)# ipv6 hold-time eigrp as-number seconds
```

（4）调整度量权重

EIGRPv6 路由度量的计算公式和 EIGRPv4 的计算公式一样：

度量值=［（*K*1×带宽）+（*K*2×带宽）÷（256－负载）+（*K*3×延迟）］×［*K*5÷（可靠性+ *K*4）］

公式中的 *K* 值称为权重（Weight），是一些常数，默认值为 *K*1=*K*3=1，*K*2=*K*4=*K*5=0，所以公式变为：

度量值=带宽+延迟；

权重参数是针对特定路由进程的，调整权重的命令是在路由进程下执行的。例如，

```
Router(config-rtr)# metric weights tos k1 k2 k3 k4 k5
```

（5）记录邻居关系变化

记录邻居关系变化可以监控路由系统的稳定性，帮助检测故障。该功能默认是开启的，如果曾经被关闭了，可以使用如下命令开启。

```
Router(config-rtr)# log-neighbor-changes
```

（6）调整邻居警告（Warning）信息

当邻居发送警告消息时，将被记录到日志里。记录警告消息的频率可以调整：

```
Router(config-rtr)# log-neighbor-warnings [seconds]
```

时间以秒为单位。

（7）关闭 Split Horizon

Split Horizon 可以避免可能出现的路由环路，但在某些网络拓扑下这个特点并不是优点，比如 NBMA 或多点 GRE（Multipoint GRE），因此需要关闭。在接口模式下，关闭它的命令是；

```
Router(config-if)#no ipv6 split-horizon eigrp as-number
```

（8）修改写入路由表的路由条目最大值

```
router(config-rtr)#maximum-paths Number of paths
```

最多 64 条。

4.3　配置路由汇总

在特定接口上配置路由汇总，汇总后的路由从该接口通告出去。汇总路由的度量值等于被汇总路由的最小度量。

```
Router(config-if)#ipv6 summary-address eigrp as-number ipv6-address [admin-distance]
```

- *admin-distance*：管理距离

例如，在 AS=1 进程中通告汇总路由 2001:0db8:0:1::/64

```
Router(config-if)# ipv6 summary-address eigrp 1 2001:0DB8:0:1::/64
```

在图 4-1 所示的例子中，路由器 P1 可以向 P2 发送总结路由，配置如下：

```
P1(config)#interface f1/1
```

```
P1(config-if)# ipv6 summary-address eigrp 88 2013:1010::/62 150
```

之后，查看 P2 的路由表：

```
P2#sh ipv6 route
IPv6 Routing Table - 8 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       …….
C    2006:1010:0:5::/64 [0/0]
      via ::, Loopback1
L    2006:1010:0:5:C806:49FF:FE46:8/128 [0/0]
      via ::, Loopback1
C    2006:1010:0:6::/64 [0/0]
      via ::, Loopback2
L    2006:1010:0:6:C806:49FF:FE46:8/128 [0/0]
      via ::, Loopback2
C    2013:127:1647:4::/64 [0/0]
      via ::, FastEthernet0/0
L    2013:127:1647:4:C806:49FF:FE46:8/128 [0/0]
      via ::, FastEthernet0/0
D    2013:1010::/62 [90/156160]
      via FE80::C803:49FF:FE21:1D, FastEthernet0/0
L    FF00::/8 [0/0]
      via ::, Null0
```

路由表显示，只接收到一条汇总路由。

再看 P1 的路由表有何变化：

```
P1#sh ipv6 route
IPv6 Routing Table - 10 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       …….
D    2006:1010:0:5::/64 [90/156160]
      via FE80::C806:49FF:FE46:8, FastEthernet1/1
D    2006:1010:0:6::/64 [90/156160]
      via FE80::C806:49FF:FE46:8, FastEthernet1/1
C    2013:127:1647:4::/64 [0/0]
      via ::, FastEthernet1/1
L    2013:127:1647:4:C803:49FF:FE21:1D/128 [0/0]
      via ::, FastEthernet1/1
D    2013:1010::/62 [150/128256]
      via ::, Null0
C    2013:1010:0:1::/64 [0/0]
      via ::, Loopback1
L    2013:1010:0:1:C803:49FF:FE21:6/128 [0/0]
```

```
        via ::, Loopback1
C     2013:1010:0:2::/64 [0/0]
        via ::, Loopback2
L     2013:1010:0:2:C803:49FF:FE21:6/128 [0/0]
        via ::, Loopback2
L     FF00::/8 [0/0]
        via ::, Null0
```

路由表显示，产生一条下一跳为 Null0、管理距离为 150 的汇总路由。

4.4　配置路由认证

路由认证功能为路由更新提供了安全机制。在路由更新消息报文中使用 MD5 对信息认证。MD5 使用密钥产生摘要字。每个密钥有自己的密钥 ID，密钥 ID 和发送消息的接口联合起来唯一地确认了 MD5 摘要。

可以使用多个密钥，每个密钥都有个声明周期（Lifetime）。系统从最小的密钥号码开始，查找有效的的第一个密钥使用。

4.4.1　配置认证

配置路由认证需以下 2 个步骤：

- 第一步定义密钥；
- 第二步在接口上启用认证并使用密钥。

使用如下命令和步骤定义密钥：

（1）定义密钥链

```
Router(config)# key chain name-of-chain
```

一个密钥链上可以有多个密钥。

（2）定义密钥

```
Router(config-keychain)# key key-id
Router(config-keychain-key)# key-string key-string-text
```

在密钥链上的多个密钥，用 key-id 编号。系统从号码小的密钥开始寻找有效的密钥使用。

（3）定义接收密钥的有效期

```
Router(config-keychain-key)#accept-lifetime start-time {infinite | end-time | duration seconds}
```

- *start-time*：密钥生效开始时间；
- infinite：永远有效；
- *end-time*：密钥有效期结束时间；
- duration：密钥有效期时长；

例如，设置某密钥从 2010 年的元月 10 号下午 14:30:00 开始生效，生效期是 7 200 秒。

```
Router(config-keychain-key)# accept-lifetime 14:30:00 Jan 10 2010 duration 7200
```

当路由器收到邻居的密钥时，如果在该时段，这个密钥就可能被使用。

（4）定义发送密钥的有效期

```
Router(config-keychain-key)# send-lifetime start-time {infinite | end-time | duration seconds}
```

路由器认证时，从有效密钥中选择使用。该命令定义的是哪些密钥可以被用作发送密钥使用。

（5）在需要启用认证的接口上启用认证

```
Router(config-if)#ipv6 authentication key-chain eigrp as-number key-china-name
```

例如：

```
P1(config-if)#ipv6 authentication key-chain eigrp 88 p1p2
```

（6）启用 MD5 认证模式

```
Router(config-if)#ipv6 authentication mode eigrp as-number md5
```

例如：

```
P1(config-if)#ipv6 authentication mode eigrp 88 md5
```

4.4.2　认证示例

以图 4-1 为例，在路由器 P1 和 P2 上有关认证的命令清单如下：

```
P1# show run
!
key chain p1p2
 key 1234
  key-string p1p2
  accept-lifetime 09:00:00 Feb 1 2013 infinite
  send-lifetime 09:00:00 Feb 1 2013 infinite
!
interface FastEthernet1/1
 no ip address
 duplex auto
 speed auto
 ipv6 address 2013:127:1647:4::/64 eui-64
 ipv6 enable
 ipv6 eigrp 88
 ipv6 authentication mode eigrp 88 md5
 ipv6 authentication key-chain eigrp 88 p1p2
 ipv6 summary-address eigrp 88 2013:1010::/62 150


P2# show run
```

```
!
key chain p1p2
 key 1234
  key-string p1p2
  accept-lifetime 09:00:00 Feb 1 2013 infinite
  send-lifetime 09:00:00 Feb 1 2013 infinite

interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 address 2013:127:1647:4::/64 eui-64
 ipv6 enable
 ipv6 eigrp 88
 ipv6 authentication mode eigrp 88 md5
 ipv6 authentication key-chain eigrp 88 p1p2
```

提示：如果 key chain 中的第一个 key 在有效 send-lifetime 和 accept-lifetime 内，但是和邻居不匹配，而第二个 key 和邻居匹配，并且在 2 个有效时间内，路由器不尝试使用第二个 key。

当认证模式不匹配时，会有如下的提示消息：

```
*Feb  1 09:54:44.623: %DUAL-5-NBRCHANGE: IPv6-EIGRP(0) 88: Neighbor FE80::C804:5FF:FE5F:1D (FastEthernet0/0) is down: authentication mode changed
```

当 key chain 更改时，有如下提示：

```
*Feb  1 09:53:32.123: %DUAL-5-NBRCHANGE: IPv6-EIGRP(0) 88: Neighbor FE80::C804:5FF:FE5F:1D (FastEthernet0/0) is down: keychain changed
```

当认证不成功时，有如下提示：

```
*Feb  1 09:54:44.943: %DUAL-5-NBRCHANGE: IPv6-EIGRP(0) 88: Neighbor FE80::C802:5FF:FE5D:8 (FastEthernet1/1) is down: Auth failure
```

4.5　本 章 小 结

EIGRPv6 的运行过程以及路由计算方法和 IPv4 下的一样，最大的不同是 EIGRP 使用新的 TLV 携带 IPv6 可达性信息（路由信息），除此之外，支持 IPv6 的 EIGRP 也有不同与 IPv4 环境下的其他特征。

EIGRP IPv6 进程直接运行在接口上，并没有类似 IPv4 下的 network 命令。只要在接口下启用了 EIGRP 协议，不需要在路由器模式下配置任何命令 EIGRP 就可以运行。

EIGRP 进程需要路由器 ID 号；EIGRPv6 进程可以“Shut Down”。

接口在 Passive 模式下不需要配置 IPv6 进程。

本章介绍了 EIGRPv6 的原理后，设计了一个简单的网络拓扑，介绍了如何配置 EIGRPv6，最后介绍了路由汇总、认证功能和优化 EIGRPv6 路由协议的参数。

第 5 章　支持 IPv6 的 OSPF

OSPF 路由协议是链路状态型路由协议，这里的链路即设备上的接口。链路状态型路由协议基于连接源和目标设备的链路状态作出路由的决定。链路状态是接口及其与邻接网络设备的的关系的描述，接口的信息即链路的信息，也就是链路的状态（信息）。这些信息包括接口的 IPv6 前缀（Prefix）、网络掩码、接口连接的网络（链路）类型、与该接口在同一网络（链路）上的路由器等信息。这些链路状态信息由不同类型的 LSA 携带，在网络上传播。

路由器把收集到的 LSA 存储在链路状态数据库中，然后运行 SPF 算法计算出路由表。链路状态数据库和路由表的本质不同在于：数据库中包含的是完整的链路状态原始数据，而路由表中列出的是到达所有已知目标网络的最短路径的列表。

OSPF 协议是为 IP 协议提供路由功能的路由协议。OSPFv2（OSPF 版本 2）支持 IPv4 的路由协议，为了让 OSPF 协议支持 IPv6，技术人员开发了 OSPFv3（OSPF 版本 3），OSPFv3 由 RFC2740 定义。

5.1　OSPFv3 与 OSPFv2

无论是 OSPFv2 还是 OSPFv3，OSPF 协议的基本运行原理是没有区别的。然而，由于 IPv4 和 IPv6 协议意义不同，地址空间大小不同，它们之间的不同之处也是必然存在的。

5.1.1　相似之处

① 路由器类型相同：包括内部路由器（Internal Router）、骨干路由器（Backbone Router）、区域边界路由器（Area Border Router）和自治系统边界路由器（Autonomous System Boundary Router）。

② 支持的区域类型相同：包括骨干区域、标准区域、末节区域、NSSA 和完全末节区域。

③ 都使用 SPF 算法。

④ DR 和 BDR 的选举过程相同。

⑤ 接口类型相同：包括点到点链路、点到多点链路、BMA 链路、NBMA 链路和虚拟链路。

⑥ LSA 的传播和老化机制相同。

⑦ 虽然 OSPFv3 的数据包携带地址空间更大的 IPv6 地址，但大多数数据包都兼容 OSPFv2。

⑧ 基本数据包类型相同，都使用 Hello、DBD、LSR、LSU 和 LSA。

⑨ 邻居的发现和邻居关系的建立机制相同。

⑩ 度量值的计算方法相同，都使用链路开销。

5.1.2　不同之处

① 术语：大多数用在 OSPFv2 中的术语都可以用来描述 OSPFv3 中相同的概念，但其中有一个术语在 OSPFv2 中和在 OSPFv3 中具有不同的含义。在 OSPFv3 中，用“链路（Link）”取代了 OSPFv2 中的“网络（Network）”或“子网（Subnet）”的概念。“接口”与“链路”相连。一条“链路”可以有多个 IP 地址，如果两个节点与同一个链路相连，即使它们不在同一个 IP 子网（IPv6 前缀）上，也能够通过该“链路”直接通话。

② OSPFv3 进程运行在链路上，而不像 OSPFv2 进程那样运行在子网上。

③ 一条链路上可以运行多个 OSPF 实例（Instance）。例如，可以使用两个实例让一条链路运行在两个区域内。

④ IPv6 地址信息仅在 LSU 的载荷（Payload）中携带。

⑤ 路由器 ID、区域 ID 和 LSA 链路状态 ID 值仍然使用 32 比特表达，因此不能使用 IPv6 地址表示。

⑥ OSPFv3 总是使用路由器 ID 来确认邻接路由器的身份。而 OSPFv2 在 NBMA 和 BMA 网络中，使用路由器 IP 地址确认邻接路由器的身份；在点到点网络和虚链路上使用路由器 ID 确认邻接路由器身份。

⑦ LSA 的传播范围（Flooding Scope）定义为如下 3 种：

- 本链路范围（Link-Local Scope）—— LSA 仅在本链路范围内传播；
- 区域范围（Area Scope）—— LSA 仅在一个区域内传播；
- AS 范围（AS scope）—— LSA 在整个 AS 内传播。

⑧ 使用本链路地址（Link-Local Address）发现邻居和完成自动配置等工作：IPv6 路由器并不转发源地址是本链路地址的数据包，OSPFv3 认为每台路由器已经为它相接的每个物理网段（物理链路）分配了本链路地址。

⑨ 认证方式不同：运行 OSPFv3 的路由器使用 IPSec 认证，因此在 OSPFv3 的报文头中不包含认证信息（参见 5.2 节）。

⑩ 数据包结构不同（参见 5.2 节）。

⑪ LSA 类型不同（参见 5.4 节）。

⑫ LSA 结构不同（参见 RFC2740）。

⑬ 一个接口可以配置多个 IPv6 地址前缀，默认地，这些前缀都被注入到 OSPFv3 中，不能人为地选择哪些前缀被注入，哪些前缀不被注入。

⑭ 如果路由器的接口上设置了 IPv4 地址，或者设置了环回接口，OSPFv3 则自动选择 IPv4 地址作为路由器 ID，否则需要为路由器设置 ID 号。

5.2　数据包结构

OSPFv3 协议直接运行在 IPv6 上，因此，OSPFv3 的数据由 IPv6 报文携带。在 IPv6 报文头的‘下一个报文头’字段中，用协议号 89 代表该数据包中携带的是 OSPFv3 的数据。

IPv6 的报文头结构如图 5-1 所示，各字段的含义如下所述。

- Version（版本号）：参数值为 6，代表 IPv6 报文头，4 比特长。
- Priority（优先级）：4 比特长。发送方可以使用该值为其发出的数据赋予优先级。
- Flow Label（流标签）：发送方可以用它标志属于同一个流的数据。
- Payload Length（载荷长度）：表示报文头后面携带的有效数据长度。
- Next Header（下一个报文头）：表示 IPv6 报文头后面的报文头类型，89 代表 OSPF 协议。
- Hops（跳数限制）：表示报文所能经由的最大路由器数。
- Source Address（源地址）：发送数据方地址。
- Destination Address（目的地地址）：接收数据方地址。

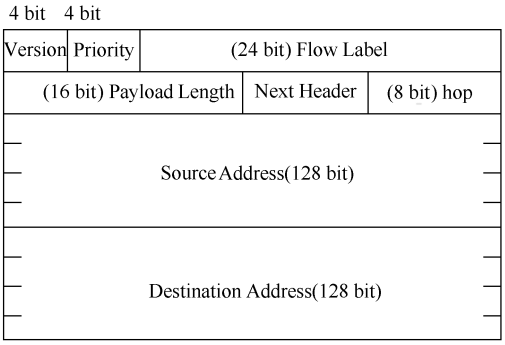

图 5-1　IPv6 报文头格式

紧跟着 IPv6 报文头的后面是 OSPFv3 的报文头，其格式如图 5-2 所示。

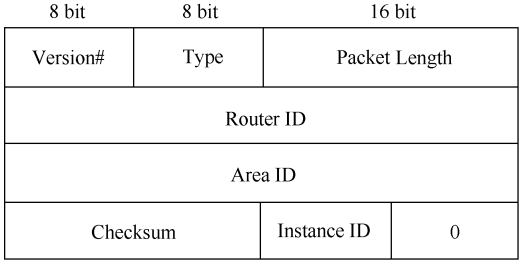

图 5-2　OSPFv3 报文头

其中，

- Version#：代表 OSPF 的版本号，其值为 3。
- Type：代表 OSPF 数据报类型。OSPF 数据报的类型包括：
  - Hello 包，类型代码为 1；
  - DDP，类型代码为 2；
  - LSR，类型代码为 3；
  - LSU，类型代码为 4；
  - LSAck，类型代码为 5。
- Packet Length：以字节为单位的 OSPF 数据报的长度，包含报文头部分。
- Router ID：发送该数据包的路由器 ID 号。
- Area ID：该数据报所属的区域。

- Checksum：整个 OSPF 数据报的校验和。
- Instance ID：实例标志号。

OSPFv3 报文头的最后 8 比特保留，值总为 0。

5.3　Hello 数据包

OSPFv3 使用两个组播地址发送各种数据包。一个是 FF02::5，代表所有运行 OSPF 协议的路由器都应该处理的数据包；另一个是 FF02::6，代表 DR 和 BDR 路由器应该处理的数据包。

Hello 数据包使用的地址是 FF02::5。Hello 数据包定期在所有运行了 OSPF 的接口上发送，用来建立和维护邻居关系。连接在同一条链路上的路由器必须对某些参数协商一致才能形成邻居关系，这些参数包含在 Hello 数据包中。Hello 数据包中也包含用来选举 DR/BDR 的参数。Hello 数据包的结构如图 5-3 所示。

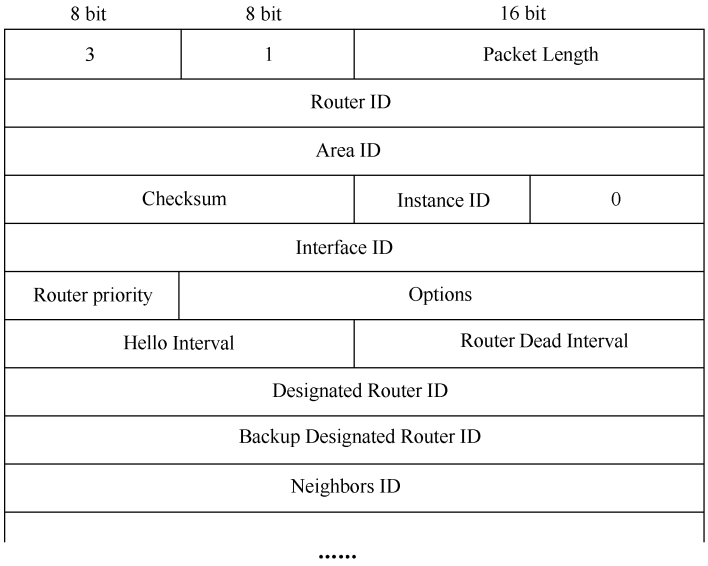

图 5-3　Hello 数据包结构

Hello 数据包中各字段的含义如下所述。

- Interface ID：接口标志符。路由器的每一个接口都有一个唯一的标志符。
- Router priority：路由器优先级。路由器根据该值选举 DR/BDR。
- Options：该 24 比特字段出现在 Hello 包、DDP 和某些 LSA 中，OSPF 路由器使用该字段实现某些与其他路由器通信的能力（详见 RFC2740）。
- Hello interval：发送 Hello 包的周期时间。
- Router Dead Interval：邻居路由器认为该路由器的失效时间。
- Designated Router ID：DR 路由器的 ID。
- Backup Designated Router ID：BDR 路由器的 ID。
- Neighbors ID：邻居列表。每个邻居 ID 占 4 字节。

5.4　IPv6 使用的 LSA 类型

5.4.1　LSA 的产生

在 IPv6 环境下生成和再次生成 LSA 的过程与 IPv4 环境下的过程时相同的。再次生成的 LSA 的序列码（LSA Sequence Number）递增，LSA 的生存时间（LSA Age）被设置为 0，LSA 被写入链路状态数据库中，并从适当的接口扩散出去。在 IPv4 环境下能够触发再次生成 LSA 的事件，同样能够触发 IPv6 环境下的路由器再次生成 LSA，除此之外，下列事件或行为也能够触发 IPv6 路由器再次生成 LSA。

- 任意一个路由器接口状态的改变。
- DR 路由器身份的改变。
- 邻居路由器从“Full”状态转变到其他状态，或者从其他状态转变到“Full”状态。
- 邻居接口 ID 的改变。
- 为链路增加了新的地址前缀，或者删除了某个地址前缀。
- 收到了新的链路 LSA（Link-LSA），并引起了连路前缀的变化。

所有的 LSA 具有共同的报文头格式，如图 5-4 所示。

LSA Age	LSA Type
Link State ID	
Advertising Router	
LAS Sequence Number	
LSA Checksum	Length

图 5-4　LSA 报文头格式

- LSA Age：已经生存的时间，以秒为单位。
- LSA Type：LSA 具体功能代码，即 LSA 类型。
- Link State ID：链路状态 ID，携带具体的链路状态信息。
- Advertising Router：通告者路由器 ID。
- LSA Sequence Number：LSA 序列码。
- LSA Checksum：LSA 校验和。
- Lenth：长度。

5.4.2　LSA 类型

OSPFv3 继承了 OSPFv2 的大多数 LSA 类型，只是名称上有所改变，并增加了新的 LSA 类型。表 5-1 列出了 OSPFv3 使用的 LSA。

表 5-1　LSA 类型一览表

LSA 名称	LSA 类型编号	LSA 类型编码
Router-LSA	1	0x2001
Network-LSA	2	0x2002
Inter-Area-Prefix-LSA	3	0x2003
Inter-Area-Router-LSA	4	0x2004
AS-External-LSA	5	0x2005
Group-Membership-LSA	6	0x2006
Type-7-LSA	7	0x2007
Link-LSA	8	0x2008
Intra-Area-Prefix-LSA	9	0x2009

① 路由器 LSA（Router-LSA）：描述在特定区域内的路由器接口的链路状态及链路开销。该类型 LSA 仅在接口所在的区域内传播。该类型 LSA 还指示出产生该类型 LSA 的路由器是否是一台 ABR 或 ASBR，以及路由器是否是虚链路的一端。类型 1LSA 也被用来通告端（Stub）网络。

② 网络 LSA（Network-LSA）：由 DR 产生，用来描述 DR 所代表的网络的链路状态信息和链路开销信息。该 LSA 是 BMA 网络上的所有链路信息和链路开销信息的集合信息。

③ 区域间前缀 LSA（Inter-Area-Prefix-LSA）：这种类型 LSA 在 OSPFv2 中叫作“概括的网络链路状态条目（Summary Net Link States）”。该类型 LSA 由 ABR 产生，用来把一个区域内的路由信息通告到另外一个区域。这种类型的 LSA 既可以描述一条具体的路由信息，也可以描述总结的路由信息。在 OSPFv3 中，地址是用前缀和前缀长度表达的，而不再使用地址和掩码的表达形式。默认路由的前缀长度用 0 代表。

④ 区域间路由器 LSA（Inter-Area-Router-LSA）：由 ASBR 产生，用以通告 ASBR 的位置。试图把数据送达外网的路由器使用该 LSA 确定通往下一跳的最佳路径。

⑤ 自治系统外 LSA（AS-External-LSA）：用来携带自治系统外路由信息的 LSA。例如，把其他 AS 的路由再发布到 OSPF 中的路由信息。在该类型的 LSA 中，地址用前缀和前缀长度表达，默认路由的前缀长度用 0 表达。

⑥ 类型 6 LSA：暂不支持。原为 MOSPF（支持组播报文的 OSPF）设计的 LSA。

⑦ 类型 7 LSA（Type-7 LSA）：由 NSSA 中的 ASBR 产生，只在 NSSA 中传播。在 ABR 上它被转换成类型 5 并且被扩散到骨干区域。

⑧ 链路 LSA（Link-LSA）：该类型 LSA 仅在本链路上传播，具有以下 3 种目的。

- 向本链路上的其他路由器提供路由器的本链路地址；
- 向本链路上的其他路由器通告 IPv6 地址前缀，使其与本链路相关联；
- 允许路由器在该类型 LSA 中插入一些选项比特，与为本链路产生的网络 LSA 相关联。

⑨ 区域内前缀 LSA（Intra-Area-Prefix-LSA）：该类型 LSA 在本区域内传播，具有以下两种功能之一。

- 通过参考网络 LSA 把 IPv6 地址前缀与转递网络（Transit Network）相关联；
- 通过参考路由器 LSA 把 IPv6 地址前缀与路由器相关联。

- 一台路由器可以为给定的区域产生多个该类型的 LSA，这些 LSA 使用链路状态 ID 字段（Link State ID）区分。DR 为它所代表的链路产生一个或多个该类型的 LSA，向整个区域内通告该链路的地址前缀。

提示：类型 6 LSA 不做讨论。

5.5　配置 OSPFv3

提示：本节使用的命令基于 IOS 12.4T。

5.5.1　基本配置命令

1. 启动 IPv6 单播路由功能

命令语法如下：

```
router(config)# ipv6 unicast-routing
```

2. 在接口上启用 IPv6

命令语法如下：

```
router(config-if)# ipv6 enable
```

该命令可以使一个没有配置 IPv6 地址的接口启用 IPv6 进程，并自动为接口配置一个本链路（Link-Local）地址。

① 下面的输出显示了 f0/0 接口上没有启用 IPv6 和启用 IPv6 后的不同情况。

```
FastEthernet0/0 is up, line protocol is up
  Hardware is AmdFE, address is 000e.38ee.39a0 (bia 000e.38ee.39a0)
  MTU 1500 bytes, BW 100000 Kbit, DLY 100 usec,
     reliability 255/255, txload 1/255, rxload 1/255
  Encapsulation ARPA, loopback not set
（略）
```

输出显示，f0/0 接口的 MAC 地址是 000e.38ee.39a0。

启用 IPv6 后的输出如下：

```
FastEthernet0/0 is up, line protocol is up
  IPv6 is enabled, link-local address is FE80::20E:38FF:FEEE:39A0
  No global unicast address is configured
  Joined group address(es):
    FF02::1
    FF02::1:FFEE:39A0
  MTU is 1500 bytes
```

```
    ICMP error messages limited to one every 100 milliseconds
    ICMP redirects are enabled
    ND DAD is enabled, number of DAD attempts: 1
    ND reachable time is 30000 milliseconds
```

输出显示，启用 IPv6 的以太接口自动把 MAC 地址转换成 EUI-64 格式，构成本链路地址。该接口也自动加入两个组播组中：本链路范围内所有 IPv6 节点组和被请求节点组中。

② 下面的输出显示的是在 s0/0 接口上启用 IPv6 后的情况。

```
Serial0/0 is up, line protocol is up
  IPv6 is enabled, link-local address is FE80::20E:38FF:FEEE:39A0
  No global unicast address is configured
  Joined group address(es):
    FF02::1
    FF02::1:FFEE:39A0
  MTU is 1500 bytes
  ICMP error messages limited to one every 100 milliseconds
  ICMP redirects are enabled
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
```

输出显示，点到点链路上的接口使用的以太接口的 MAC 地址转换为 EUI-64 地址，构成其本链路地址。

3. 为接口设置单点传送地址

为运行 IPv6 的路由器接口设置地址的方法有多种，下面介绍 3 种常用的方法。

① 直接手工设置地址。

命令语法如下：

```
router(config-if)# ipv6 address ipv6-address/prefix-length
```

ipv6-address/prefix-length：IPv6 地址及地址前缀长度。

② 自动使用基于 EUI-64 格式的主机地址。

命令语法如下：

```
router(config-if)# ipv6 address ipv6-prefix/prefix-length eui-64
```

- *ipv6-prefix/prefix-length*：IPv6 地址前缀及前缀长度。

该命令相当于为接口设置了网络前缀，节点号就直接使用了符合 EUI-64 格式的接口地址（接口物理地址）。如果设置的前缀长度大于 64，前缀中的比特位优先于接口地址的比特位。

在接口上设置 IPv6 地址，就自动启用了该接口的 IPv6 进程，并且“no ipv6 enable”命令不能禁止该接口上的 IPv6 进程。

例如，为接口 ethernet 0 设置 IPv6 地址：

```
Router(config)# interface fastethernet 0/0
Router(config-if)# ipv6 address 2006 : 1010 : 0 : 1 :: /64    eui-64
```

设置了全球单点传送地址后的输出如下：

```
FastEthernet0/0 is up, line protocol is up
  IPv6 is enabled, link-local address is FE80::20E:38FF:FEEE:39A0
  Global unicast address(es):
    2006:1010:0:1:20E:38FF:FEEE:39A0, subnet is 2006:1010:0:1::/64
  Joined group address(es):
    FF02::1
    FF02::1:FFEE:39A0
  MTU is 1500 bytes
  ICMP error messages limited to one every 100 milliseconds
  ICMP redirects are enabled
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
```

③ 自动配置地址。

命令语法如下：

```
Router(config-if)# ipv6 address autoconfig
```

该命令启动了 IPv6 的自动地址配置功能，自动发现链路上的地址前缀，然后在前缀后面自动加上 EUI-64 格式的接口物理地址，构成接口的 IPv6 地址。

4. 设置路由器 ID

在 IPv6 环境下，可以使用如下几种方式为路由器设置 ID 号：

① 如果物理接口上设置了 IPv4 地址，路由器自动选择活跃接口上的最高 IPv4 地址作为其 ID 号。

② 如果设置了 Loopback 接口，则优先使用 Loopback 接口地址作为路由器 ID；如果存在多个 Loopback 接口，则使用最高接口地址。

③ 如果没有任何接口配置了 IPv4 地址，则必须使用如下命令为路由器设置 ID 号：

```
router(config)# ipv6 router ospf process-id
router(config-rtr)# router-id    ipv4-format-id
```

- *process-id*：进程号；
- *ipv4-format-id*：IPv4 格式的路由器 ID 号。

5. 在接口上启用 OSPF

命令语法如下：

```
Router(config-if)# ipv6 ospf process-id area area-id [instance instance-id]
```

- *process-id*：进程号；
- *area-id*：区域号；
- *instance-id*：实例号。

该命令把接口置于某个 OSPF 进程的某个区域中，并且还可以声明其使用的实例编号。

在接口上启用 OSPF，可以自动启用 OSPF 进程，不需要再单独使用令 **ipv6 router ospf** *process-id* 启用进程；该命令也同时启用了 IPv6 的单播路由功能。

6. 配置路由总结

命令语法如下：

```
Router(config)# ipv6 router ospf process-id
Router(config-rtr)# area area-id range ipv6-prefix/prefix-length [advertise | not-advertise] [cost cost]
```

该命令的含义和 OSPFv2 中的路由总结命令含义相同，其中，

- *area-id*：表示总结某个区域内的路由；
- *ipv6-prefix/prefix-length*：总结后的路由前缀及前缀长度；
- advertise：通告该总结路由条目；
- not-advertise：不通告该总结路由条目；
- *cost*：为该总结路由赋予的开销值，默认地，总结路由条目的开销使用被总结路由中的最高开销值。

7. OSPFv3 配置示例

提示：下面的实验环境使用的 IOS 版本是 12.3(20)，高级企业特性集。

配置示例中使用的拓扑如图 5-5 所示。路由器 R1 和 R2 通过点到点链路把两个局域网连接起来。两个串行接口运行在区域 0 中。R1 的以太接口和环回接口运行在区域 3 中，R2 的以太接口和环回接口运行在区域 2 中。下面分别配置 R1 和 R2。

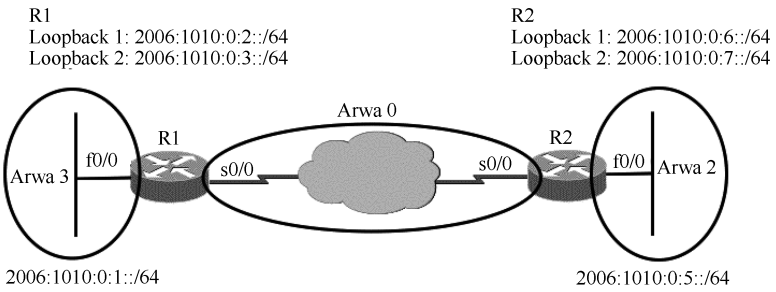

图 5-5　配置示例拓扑

（1）配置路由器 R1

```
Router1#config t
Router1(config)#ipv6 unicast-routing
router1(config)#ipv6 router ospf 1
router1(config-rtr)#router-id 1.1.1.1
router1(config-rtr)#exit
router1(config)#interface f0/0
router1(config-if)#ipv6 address 2006:1010:0:1::/64 eui-64
router1(config-if)# ipv6 ospf 1 area 3
router1(config-if)#interface s0/0
router1(config-if)# ipv6 ospf 1 area 0
```

```
router1(config-if)#exit
router1(config)#interface loopback 1
router1(config-if)#ipv6 address 2006:1010:0:2::/64 eui-64
router1(config-if)#ipv6 ospf 1 area 3
router1(config-if)#interface loopback 2
router1(config-if)#ipv6 add 2006:1010:0:3::/64 eui-64
router1(config-if)#ipv6 ospf 1 area 3
router1(config-if)#exit
router1(config)#
```

配置完成后的配置列表如下所示（仅列出相关内容）：

```
hostname router1
!
ipv6 unicast-routing
!
interface Loopback1
 no ip address
 ipv6 address 2006:1010:0:2::/64 eui-64
 ipv6 ospf 1 area 3
!
interface Loopback2
 no ip address
 ipv6 address 2006:1010:0:3::/64 eui-64
 ipv6 ospf 1 area 3
!
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 address 2006:1010:0:1::/64 eui-64
 ipv6 enable
 ipv6 ospf 1 area 3
!
interface Serial0/0
 no ip address
 ipv6 enable
 ipv6 ospf 1 area 0
 no fair-queue
!
ipv6 router ospf 1
 router-id 1.1.1.1
 log-adjacency-changes
```

（2）配置路由器 R2

```
Router2#config t
Router2(config)#ipv6 unicast-routing
Router2(config)#ipv6 router ospf 1
Router2(config-rtr)#router-id 2.2.2.2
Router2(config-rtr)#exit
Router2(config)#interface f0/0
Router2(config-if)#ipv6 address 2006:1010:0:5::/64 eui-64
Router2(config-if)# ipv6 ospf 1 area 2
Router2(config-if)#interface s0/0
Router2(config-if)# ipv6 ospf 1 area 0
Router2(config-if)#exit
Router2(config)#interface loopback 1
Router2(config-if)#ipv6 address 2006:1010:0:6::/64 eui-64
Router2(config-if)#ipv6 ospf 1 area 2
Router2(config-if)#interface loopback 2
Router2(config-if)#ipv6 add 2006:1010:0:7::/64 eui-64
Router2(config-if)#ipv6 ospf 1 area 2
Router2(config-if)#exit
Router2(config)#
```

配置完成后的配置列表如下所示（仅列出相关内容）：

```
hostname router2
!
ipv6 unicast-routing
!
interface Loopback1
 no ip address
 ipv6 address 2006:1010:0:6::/64 eui-64
 ipv6 ospf 1 area 2
!
interface Loopback2
 no ip address
 ipv6 address 2006:1010:0:7::/64 eui-64
 ipv6 ospf 1 area 2
!
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 address 2006:1010:0:5::/64 eui-64
 ipv6 ospf 1 area 2
```

```
!
interface Serial0/0
 no ip address
 ipv6 enable
 ipv6 ospf 1 area 0
!
ipv6 router ospf 1
 router-id 2.2.2.2
```

5.5.2　检查 OSPFv3 的运行状态

1. show ipv6 ospf

```
router2#sh ipv6 ospf
 Routing Process "ospfv3 1" with ID 2.2.2.2
 It is an area border router
 SPF schedule delay 5 secs, Hold time between two SPFs 10 secs
 Minimum LSA interval 5 secs. Minimum LSA arrival 1 secs
 LSA group pacing timer 240 secs
 Interface flood pacing timer 33 msecs
 Retransmission pacing timer 66 msecs
 Number of external LSA 0. Checksum Sum 0x000000
 Number of areas in this router is 2. 2 normal 0 stub 0 nssa
    Area BACKBONE(0)
        Number of interfaces in this area is 1
        SPF algorithm executed 7 times
        Number of LSA 8. Checksum Sum 0x04C5E5
        Number of DCbitless LSA 0
        Number of indication LSA 0
        Number of DoNotAge LSA 0
        Flood list length 0
    Area 2
        Number of interfaces in this area is 1
        SPF algorithm executed 3 times
        Number of LSA 6. Checksum Sum 0x013219
        Number of DCbitless LSA 0
        Number of indication LSA 0
        Number of DoNotAge LSA 0
        Flood list length 0
```

该命令显示了路由器 R2 上 OSPFv3 的基本运行情况。

该路由器上运行的 OSPFv3 的进程号为 1，路由器 ID 号是 2.2.2.2，该路由器是一台区域边

界路由器。该路由器上有两个区域，并且都是标准区域（Normal Area）。

在骨干（Backbone）区域内有一个接口，该区域中的 SPF 算法已经执行了 7 次，区域 0 中的 LSA 共有 8 个。

在区域 2 中有一个接口，该区域中的 SPF 算法已经执行了 3 次；该区域中的 LSA 共有 6 个。

2. show ipv6 ospf neighbor

```
router1#sh ipv6 ospf neighbor

Neighbor ID     Pri   State           Dead Time   Interface ID    Interface
2.2.2.2           1   FULL/  -        00:00:34    5               Serial0/0
router1#
```

该命令显示了路由器 R1 的邻居情况，其中，Interface ID 一列显示的是邻居路由器与路由器 R1 相连接口的 ID 号。

3. show ipv6 ospf neighbor detail

```
router1#sh ipv6 ospf neighbor detail
 Neighbor 2.2.2.2
    In the area 0 via interface Serial0/0
    Neighbor: interface-id 5, link-local address FE80::20C:31FF:FEE9:860
    Neighbor priority is 1, State is FULL, 6 state changes
    Options is 0x831C1F25
    Dead timer due in 00:00:32
    Neighbor is up for 00:13:28
    Index 1/1/1, retransmission queue length 0, number of retransmission 1
    First 0x0(0)/0x0(0)/0x0(0) Next 0x0(0)/0x0(0)/0x0(0)
    Last retransmission scan length is 2, maximum is 2
    Last retransmission scan time is 0 msec, maximum is 0 msec
```

该命令报告了邻居的详细信息。路由器 R1 在区域 0 中的 s0/0 接口上建立了一个邻居关系，邻居的 ID 为 2.2.2.2，邻居的接口 ID 为 5，该接口的本链路地址为 **FE80::20C:31FF:FEE9:860**。邻居的优先级为 1，处于数据库同步状态（Full），曾经有过 6 次状态改变。

4. show ipv6 ospf interface

```
router2#sh ipv6 ospf int s0/0
① Serial0/0 is up, line protocol is up
② Link Local Address FE80::20C:31FF:FEE9:860, Interface ID 5
③ Area 0, Process ID 1, Instance ID 0, Router ID 2.2.2.2
④ Network Type POINT_TO_POINT, Cost: 64
  Transmit Delay is 1 sec, State POINT_TO_POINT,
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    Hello due in 00:00:06
```

```
    Index 1/1/1, flood queue length 0
    Next 0x0(0)/0x0(0)/0x0(0)
    Last flood scan length is 2, maximum is 2
    Last flood scan time is 0 msec, maximum is 0 msec
⑤ Neighbor Count is 1, Adjacent neighbor count is 1
⑥ Adjacent with neighbor 1.1.1.1
    Suppress hello for 0 neighbor(s)
```

- 第一行报告了接口是否处于 Up 状态；
- 第二行报告了接口的本链路地址和接口 ID；
- 第三行报告了接口运行在的区域号、进程号、实例号及路由器 ID 号；
- 第四行报告了接口运行在点到点网络类型上；
- 第五行报告了在该接口上物理邻居数量和相邻关系数量；
- 第六行报告了与邻居 1.1.1.1 建立了相邻关系。

5. show ipv6 ospf database

下面显示的是在路由器 R2 的两个环回接口还没有运行在 OSPF 进程中时，路由器 R1 上的链路状态数据库。

```
router1#sh ipv6 ospf database

            OSPFv3 Router with ID (1.1.1.1) (Process ID 1)

                Router Link States (Area 0)

ADV Router      Age         Seq#       Fragment ID   Link count   Bits
1.1.1.1         880         0x80000002 0             1            B
2.2.2.2         885         0x80000005 0             1            B

                Inter Area Prefix Link States (Area 0)

ADV Router      Age         Seq#       Prefix
1.1.1.1         875         0x80000001 2006:1010:0:1::/64
1.1.1.1         253         0x80000001 2006:1010:0:2:20E:38FF:FEEE:39A0/128
1.1.1.1         229         0x80000001 2006:1010:0:3:20E:38FF:FEEE:39A0/128
2.2.2.2         1699        0x80000001 2006:1010:0:5::/64

                Link (Type-8) Link States (Area 0)

ADV Router      Age         Seq#       Link ID    Interface
1.1.1.1         880         0x80000002 5          Se0/0
2.2.2.2         1754        0x80000001 5          Se0/0
```

```
                Router Link States (Area 3)

ADV Router      Age         Seq#         Fragment ID   Link count   Bits
1.1.1.1         233         0x80000004   0             0            B

                Inter Area Prefix Link States (Area 3)

ADV Router      Age         Seq#         Prefix
1.1.1.1         877         0x80000001   2006:1010:0:5::/64

                Link (Type-8) Link States (Area 3)

ADV Router      Age         Seq#         Link ID      Interface
1.1.1.1         887         0x80000001   4            Fa0/0

                Intra Area Prefix Link States (Area 3)

ADV Router      Age         Seq#         Link ID      Ref-lstype   Ref-LSID
1.1.1.1         235         0x80000003   0            0x2001       0
router1#
```

其中，

- “ADV Router”：此列表示通告 LSA 的路由器 ID；
- “Age”：此列表示 LSA 在数据库中存留的时间；
- “Seq#”：此列表示 LSA 的序列码；
- “Fragment ID”：此列表示 LSA 的分段编号（如果 LSA 过大会分段传送）；
- “Link count”：此列表示路由器在该区域内的链路数量；
- “Bits”：此列的值为“B”，表示是一台区域边界路由器；
- “Prefix”：此列表示 LSA 携带的地址前缀；
- “Link ID”：此列表示接口的 ID 号；
- “interface”：此列表示路由器在哪些接口上产生的类型 8 LSA；
- “Ref-lstype”：此列表示参考的是哪种类型的 LSA；
- “Ref-LSID”：此列表示参考的链路 ID 号。

提示：相邻关系没有达到 Full 状态或者处于 Down 状态的接口，以及环回接口不被描述，因此，路由器 R1 在区域 3 中的接口数量为 0。

6. show ipv6 ospf database router

```
router1#sh ipv6 ospf database router

            OSPFv3 Router with ID (1.1.1.1) (Process ID 1)
```

Router Link States (Area 0)

1. LS age: 911
2. Options: (V6-Bit E-Bit R-bit DC-Bit)
3. LS Type: Router Links
4. Link State ID: 0
5. Advertising Router: 1.1.1.1
6. LS Seq Number: 80000002
7. Checksum: 0xC0E7
8. Length: 40
9. Area Border Router
10. Number of Links: 1

11. Link connected to: another Router (point-to-point)
12. Link Metric: 64
13. Local Interface ID: 5
14. Neighbor Interface ID: 5
15. Neighbor Router ID: 2.2.2.2

Routing Bit Set on this LSA
LS age: 917
Options: (V6-Bit E-Bit R-bit DC-Bit)
LS Type: Router Links
Link State ID: 0
Advertising Router: 2.2.2.2
LS Seq Number: 80000005
Checksum: 0x4A5B
Length: 40
Area Border Router
Number of Links: 1

Link connected to: another Router (point-to-point)
Link Metric: 64
Local Interface ID: 5
Neighbor Interface ID: 5
Neighbor Router ID: 1.1.1.1

Router Link States (Area 3)

```
    LS age: 268
    Options: (V6-Bit E-Bit R-bit DC-Bit)
    LS Type: Router Links
    Link State ID: 0
    Advertising Router: 1.1.1.1
    LS Seq Number: 80000004
    Checksum: 0xCD3C
    Length: 24
    Area Border Router
    Number of Links: 0
```

下面只解释路由器 R1 在区域 0 中通告的路由器 LSA（LSA 类型 1），其他部分的含义请读者参考该部分的解释去理解。

- 第一行：该 LSA 在数据库中存活时间；
- 第二行：选项比特的设置情况，有关选项比特的含义和功能意义深刻，已超出本书讨论范围，请读者参考 RFC2740；
- 第三行：LSA 类型；
- 第四行：LSA ID；
- 第五行：通告者路由器 ID；
- 第六行：LSA 的序列码；
- 第七行：校验值；
- 第八行：长度；
- 第九行：该路由器是一台 ABR；
- 第十行：该路由器有一条链路在区域 0 中；
- 第十一行：该链路与另一台路由器相连，是一条点到点链路；
- 第十二行：链路度量值；
- 第十三行：本路由器接口 ID 号；
- 第十四行：邻居路由器接口 ID 号；
- 第十五行：邻居路由器 ID 号；

7. show ipv6 ospf database inter-area prefix

```
router1#sh ipv6 ospf database inter-area prefix

            OSPFv3 Router with ID (1.1.1.1) (Process ID 1)

                Inter Area Prefix Link States (Area 0)

    LS age: 1296
    LS Type: Inter Area Prefix Links
    Link State ID: 0
```

```
Advertising Router: 1.1.1.1
LS Seq Number: 80000001
Checksum: 0xCEDB
Length: 36
Metric: 1
Prefix Address: 2006:1010:0:1::
Prefix Length: 64, Options: None

LS age: 673
LS Type: Inter Area Prefix Links
Link State ID: 1
Advertising Router: 1.1.1.1
LS Seq Number: 80000001
Checksum: 0xB49C
Length: 44
Metric: 0
Prefix Address: 2006:1010:0:2:20E:38FF:FEEE:39A0
Prefix Length: 128, Options: None

LS age: 653
LS Type: Inter Area Prefix Links
Link State ID: 2
Advertising Router: 1.1.1.1
LS Seq Number: 80000001
Checksum: 0xBC92
Length: 44
Metric: 0
Prefix Address: 2006:1010:0:3:20E:38FF:FEEE:39A0
Prefix Length: 128, Options: None

Routing Bit Set on this LSA
LS age: 140
LS Type: Inter Area Prefix Links
Link State ID: 0
Advertising Router: 2.2.2.2
LS Seq Number: 80000002
Checksum: 0xF6AA
Length: 36
Metric: 1
Prefix Address: 2006:1010:0:5::
Prefix Length: 64, Options: None
```

```
                Inter Area Prefix Link States (Area 3)

1. LS age: 1302
2. LS Type: Inter Area Prefix Links
3. Link State ID: 0
4. Advertising Router: 1.1.1.1
5. LS Seq Number: 80000001
6. Checksum: 0x98CD
7. Length: 36
8. Metric: 65
9. Prefix Address: 2006:1010:0:5::
10. Prefix Length: 64, Options: None

router1#
```

编号部分是路由器 R1 向区域 3 通告的区域间 LSA（LSA 类型 3）。

- 第一行：该 LSA 在数据库中存活时间；
- 第二行：LSA 类型；
- 第三行：LSA ID；
- 第四行：通告者路由器 ID；
- 第五行：LSA 的序列码；
- 第六行：校验值；
- 第七行：LSA 长度；
- 第八行：路由前缀的度量值；
- 第九行：通告的路由前缀；
- 第十行：前缀长度及选项参数。

8. show ipv6 ospf database prefix

```
router1#sh ipv6 ospf database prefix

            OSPFv3 Router with ID (1.1.1.1) (Process ID 1)

                Intra Area Prefix Link States (Area 3)

  Routing Bit Set on this LSA
  LS age: 700
  LS Type: Intra-Area-Prefix-LSA
  Link State ID: 0
  Advertising Router: 1.1.1.1
  LS Seq Number: 80000003
```

```
Checksum: 0x840F
Length: 84
Referenced LSA Type: 2001
Referenced Link State ID: 0
Referenced Advertising Router: 1.1.1.1
Number of Prefixes: 3
Prefix Address: 2006:1010:0:3:20E:38FF:FEEE:39A0
Prefix Length: 128, Options: LA , Metric: 0
Prefix Address: 2006:1010:0:2:20E:38FF:FEEE:39A0
Prefix Length: 128, Options: LA , Metric: 0
Prefix Address: 2006:1010:0:1::
Prefix Length: 64, Options: None, Metric: 1
```

以上是路由器 R1 向区域 3 中通告的 LSA 类型 9。下面只解释该类型 LSA 携带的专有信息。

- Referenced LSA Type: 2001：表示参考的 LSA 类型。
- Referenced Link State ID: 0：表示参考的链路 ID，如果参考的 LSA 类型为 1，该值为 0。
- Referenced Advertising Router: 1.1.1.1：表示被参考的路由器 ID。
- Number of Prefixes: 3：表示该 LSA 携带了 3 个前缀，这 3 个前缀分别是
  - 2006: 1010:0:3:20E:38FF:FEEE:39A0，前缀长度为 128，度量为 0，选项 LA（Local Address）表示该前缀实际上是通告者路由器的接口地址；
  - 2006:1010:0:2:20E:38FF:FEEE:39A0，前缀长度为 128，度量为 0，选项 LA（Local Address）表示该前缀是通告者路由器的接口地址；
  - 2006:1010:0:1::，前缀长度是 64，度量为 1。

9. show ipv6 route

下面是路由器 R2 上的 3 个环回接口运行在 OSPF 进程后，两台路由器上的路由表。

（1）路由器 R1 的路由表

```
router1#sh ipv6 route
IPv6 Routing Table - 11 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       U - Per-user Static route
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
C    2006:1010:0:1::/64 [0/0]
      via ::, FastEthernet0/0
L    2006:1010:0:1:20E:38FF:FEEE:39A0/128 [0/0]
      via ::, FastEthernet0/0
C    2006:1010:0:2::/64 [0/0]
      via ::, Loopback1
L    2006:1010:0:2:20E:38FF:FEEE:39A0/128 [0/0]
```

```
        via ::, Loopback1
C     2006:1010:0:3::/64 [0/0]
        via ::, Loopback2
L     2006:1010:0:3:20E:38FF:FEEE:39A0/128 [0/0]
        via ::, Loopback2
OI    2006:1010:0:5::/64 [110/65]
        via FE80::20C:31FF:FEE9:860, Serial0/0
OI    2006:1010:0:6:20C:31FF:FEE9:860/128 [110/64]
        via FE80::20C:31FF:FEE9:860, Serial0/0
OI    2006:1010:0:7:20C:31FF:FEE9:860/128 [110/64]
        via FE80::20C:31FF:FEE9:860, Serial0/0
L     FE80::/10 [0/0]
        via ::, Null0
L     FF00::/8 [0/0]
        via ::, Null0
router1#
```

（2）路由器 R2 的路由表

```
router2#sh ipv6 route
IPv6 Routing Table - 11 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
           U - Per-user Static route
           I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
           O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
           ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
OI    2006:1010:0:1::/64 [110/65]
        via FE80::20E:38FF:FEEE:39A0, Serial0/0
OI    2006:1010:0:2:20E:38FF:FEEE:39A0/128 [110/64]
        via FE80::20E:38FF:FEEE:39A0, Serial0/0
OI    2006:1010:0:3:20E:38FF:FEEE:39A0/128 [110/64]
        via FE80::20E:38FF:FEEE:39A0, Serial0/0
C     2006:1010:0:5::/64 [0/0]
        via ::, FastEthernet0/0
L     2006:1010:0:5:20C:31FF:FEE9:860/128 [0/0]
        via ::, FastEthernet0/0
C     2006:1010:0:6::/64 [0/0]
        via ::, Loopback1
L     2006:1010:0:6:20C:31FF:FEE9:860/128 [0/0]
        via ::, Loopback1
C     2006:1010:0:7::/64 [0/0]
        via ::, Loopback2
```

```
L     2006:1010:0:7:20C:31FF:FEE9:860/128 [0/0]
        via ::, Loopback2
L     FE80::/10 [0/0]
        via ::, Null0
L     FF00::/8 [0/0]
        via ::, Null0
router2#
```

5.5.3　NBMA 配置示例

拓扑结构如图 5-6 所示，路由器 R1 和路由器 R2 通过帧中继相连。两台路由器都运行在 NBMA 模式，因此需要手工指定邻居，也需要手工做 IPv6 地址和 DLCI 值的映射。手工配置邻居和地址影射都使用邻居的本链路地址。R1 使用的 DLCI 值是 110；R2 使用的 DLCI 值是 120。

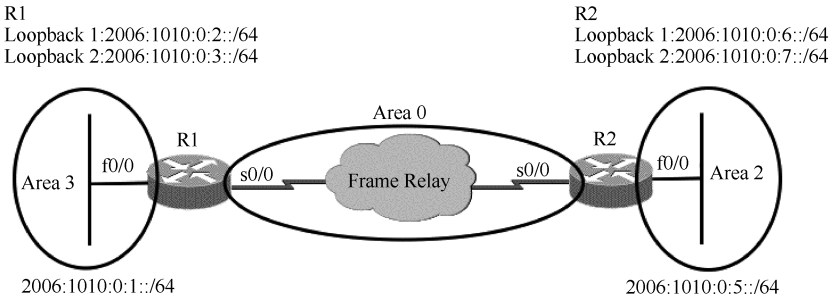

图 5-6　NBMA 配置拓扑图

（1）路由器 R1 的配置列表（只列出相关配置）

```
hostname router1
!
interface Loopback1
 no ip address
 ipv6 address 2006:1010:0:2::/64 eui-64
 ipv6 ospf 1 area 3
!
interface Loopback2
 no ip address
 ipv6 address 2006:1010:0:3::/64 eui-64
 ipv6 ospf 1 area 3
!
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
```

```
 ipv6 address 2006:1010:0:1::/64 eui-64
 ipv6 enable
 ipv6 ospf 1 area 3
!
interface Serial0/0
 no ip address
 encapsulation frame-relay IETF
 ipv6 enable
 ipv6 ospf neighbor FE80::20C:31FF:FEE9:860 priority 9
 ipv6 ospf 1 area 0
 frame-relay map ipv6 FE80::20C:31FF:FEE9:860 110 IETF
!
ipv6 router ospf 1
 router-id 1.1.1.1
```

在 R1 上手工指定邻居时使用了优先级（Priority）参数，使邻居成为了 DR。

（2）路由器 R2 的配置列表（只列出相关配置）

```
hostname router2
!
interface Loopback1
 no ip address
 ipv6 address 2006:1010:0:6::/64 eui-64
 ipv6 ospf 1 area 2
!
interface Loopback2
 no ip address
 ipv6 address 2006:1010:0:7::/64 eui-64
 ipv6 ospf 1 area 2
!
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 address 2006:1010:0:5::/64 eui-64
 ipv6 ospf 1 area 2
 no keepalive
!
interface Serial0/0
 no ip address
 encapsulation frame-relay IETF
 ipv6 enable
```

```
 ipv6 ospf neighbor FE80::20E:38FF:FEEE:39A0
 ipv6 ospf 1 area 0
 frame-relay map ipv6 FE80::20E:38FF:FEEE:39A0 120 IETF
!
ipv6 router ospf 1
 router-id 2.2.2.2
```

（3）路由器 R1 的邻居数据库

```
router1# sh ipv6 ospf neighbor detail
 Neighbor 2.2.2.2
     In the area 0 via interface Serial0/0
     Neighbor: interface-id 5, link-local address FE80::20C:31FF:FEE9:860
     Neighbor priority is 1 (configured 9), State is FULL, 12 state changes
     DR is 1.1.1.1 BDR is 2.2.2.2
（略）
```

（4）路由器 R2 的接口参数

```
router2#sh ipv6 ospf interface s0/0
Serial0/0 is up, line protocol is up
   Link Local Address FE80::20C:31FF:FEE9:860, Interface ID 5
   Area 0, Process ID 1, Instance ID 0, Router ID 2.2.2.2
   Network Type NON_BROADCAST, Cost: 64
   Transmit Delay is 1 sec, State DR, Priority 1
   Designated Router (ID) 2.2.2.2, local address FE80::20C:31FF:FEE9:860
   Backup Designated router (ID) 1.1.1.1, local address FE80::20E:38FF:FEEE:39A0
   Timer intervals configured, Hello 30, Dead 120, Wait 120, Retransmit 5
     Hello due in 00:00:28
   Index 1/1/1, flood queue length 0
   Next 0x0(0)/0x0(0)/0x0(0)
   Last flood scan length is 1, maximum is 2
   Last flood scan time is 0 msec, maximum is 0 msec
   Neighbor Count is 1, Adjacent neighbor count is 1
     Adjacent with neighbor 1.1.1.1    (Backup Designated Router)
   Suppress hello for 0 neighbor(s)
```

（5）链路状态数据库

```
router2#sh ipv6 ospf database

                OSPFv3 Router with ID (2.2.2.2) (Process ID 1)

                    Router Link States (Area 0)
```

```
ADV Router      Age         Seq#         Fragment ID   Link count    Bits
1.1.1.1         159         0x8000000C   0             1             B
2.2.2.2         141         0x8000000A   0             1             B

                Net Link States (Area 0)

ADV Router      Age         Seq#         Link ID       Rtr count
2.2.2.2         188         0x80000001   5             2        （NBMA 链路上的路由器数量）

                Inter Area Prefix Link States (Area 0)

ADV Router      Age         Seq#         Prefix
1.1.1.1         380         0x80000002   2006:1010:0:3:20E:38FF:FEEE:39A0/128
1.1.1.1         380         0x80000002   2006:1010:0:2:20E:38FF:FEEE:39A0/128
1.1.1.1         380         0x80000002   2006:1010:0:1::/64
2.2.2.2         989         0x80000001   2006:1010:0:7:20C:31FF:FEE9:860/128
2.2.2.2         989         0x80000001   2006:1010:0:6:20C:31FF:FEE9:860/128
2.2.2.2         124         0x80000004   2006:1010:0:5::/64

                Link (Type-8) Link States (Area 0)

ADV Router      Age         Seq#         Link ID       Interface
1.1.1.1         386         0x80000002   5             Se0/0
2.2.2.2         1005        0x80000001   5             Se0/0

                Router Link States (Area 2)

ADV Router      Age         Seq#         Fragment ID   Link count    Bits
2.2.2.2         965         0x80000002   0             0             B

                Inter Area Prefix Link States (Area 2)

ADV Router      Age         Seq#         Prefix
2.2.2.2         180         0x80000001   2006:1010:0:1::/64
2.2.2.2         180         0x80000001   2006:1010:0:2:20E:38FF:FEEE:39A0/128
2.2.2.2         180         0x80000001   2006:1010:0:3:20E:38FF:FEEE:39A0/128

                Link (Type-8) Link States (Area 2)

ADV Router      Age         Seq#         Link ID       Interface
```

```
2.2.2.2          1005          0x80000001    4             Fa0/0

                Intra Area Prefix Link States (Area 2)

ADV Router       Age           Seq#          Link ID       Ref-lstype    Ref-LSID
2.2.2.2          1005          0x80000001    0             0x2001        0
router2#
```

（6）类型 2 LSA 的细节

```
router2#sh ipv6 ospf database network

            OSPFv3 Router with ID (2.2.2.2) (Process ID 1)

                Net Link States (Area 0)

  LS age: 217
  Options: (V6-Bit E-Bit R-bit DC-Bit)
  LS Type: Network Links
  Link State ID: 5 (Interface ID of Designated Router)
  Advertising Router: 2.2.2.2
  LS Seq Number: 80000001
  Checksum: 0xF8F6
  Length: 32
        Attached Router: 2.2.2.2
        Attached Router: 1.1.1.1
```

5.5.4　路由总结

设置路由总结的语法如下：

```
router(config-rtr)#area area-id range ipv6-prefix / prefix-length [advertise | not-advertise] [cost cost]
```

- *area-id*：被总结路由所在的区域号；
- *ipv6-prefix / prefix-length*：总结后的路由前缀及前缀长度；
- advertise：通告该总结路由；
- not-advertise：不向邻居通告该总结路由；
- *cost*：为总结路由人为赋的开销值。

例如，在上述 NBMA 的环境中，在路由器 R1 上对区域 3 的路由进行总结，把路由的前缀总结到 62 位长，并赋予该总结路由的开销为 99，命令如下：

router1(config-rtr)#area 3 range 2006:1010::/62 cost 99

之后，在路由器 R2 上观察到的路由表如下：

```
router2#sh ipv6 route
IPv6 Routing Table - 9 entries
```

```
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       U - Per-user Static route
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
OI  2006:1010::/62 [110/163]
     via FE80::20E:38FF:FEEE:39A0, Serial0/0
C   2006:1010:0:5::/64 [0/0]
     via ::, FastEthernet0/0
L   2006:1010:0:5:20C:31FF:FEE9:860/128 [0/0]
     via ::, FastEthernet0/0
C   2006:1010:0:6::/64 [0/0]
     via ::, Loopback1
L   2006:1010:0:6:20C:31FF:FEE9:860/128 [0/0]
     via ::, Loopback1
C   2006:1010:0:7::/64 [0/0]
     via ::, Loopback2
L   2006:1010:0:7:20C:31FF:FEE9:860/128 [0/0]
     via ::, Loopback2
L   FE80::/10 [0/0]
     via ::, Null0
L   FF00::/8 [0/0]
     via ::, Null0
```

提示：在 R1 的路由表中将生成一条如下的路由——　O　　2006:1010::/62 [110/0] via ::, Null0，目的是为避免环路。

5.5.5　配置 OSPFv3 认证

与 OSPFv2 相同，OSPFv3 也可以支持两种认证方式：接口认证方式和区域认证方式。与 OSPFv2 不同的是，OSPFv3 使用 IPSec 认证。

1. 启用接口认证

语法如下：

```
router(config-if)# ipv6 ospf authentication ipsec spi spi {md5 | sha1} [key-encryption-type {key | null}]
```

- *spi*：安全策略索引（Security Policy Index）值，合法的十进制值为 256～4 294 967 295。
- *key-encryption-type*：密钥类型。0（默认）——未加密的密钥值；7——已加密的密钥值。
- *key*：计算 MD5 值使用的密钥值。32 位十六进制数。
- null：覆盖区域认证的参数。如果同时启用了区域认证和接口认证，需要使用该参数；如果没有启用区域认证可以不使用该参数。

提示：启用认证的接口必须使用相同的策略，即相同的 spi 值和 key 值。

例如，图 5-7 所示的两台路由器启用了接口认证，具体配置如下（只列出认证部分）：

图 5-7　接口认证示例图

（1）配置 R2

```
router2(config)# interface s0/1
Router2(config-if)# ipv6 ospf authentication ipsec spi 800 md5 1234567890abcdef0123456789abcdef
Router2(config-if)# ipv6 ospf authentication null
```

（2）配置 R3

```
router3(config)# interface s0/1
Router3(config-if)# ipv6 ospf authentication ipsec spi 800 md5 1234567890abcdef0123456789abcdef
Router3(config-if)# ipv6 ospf authentication null
```

2. 启用区域认证

启用区域认证意味着区域内的所有接口都必须经过认证，因此，区域内的路由器都需要启用认证。启用区域认证的语法如下：

```
router(config-rtr)#area area-id authentication ipsec spi spi {md5|sha1} [key-encryption-type] key
```

- *area-id*：启用认证的区域号；
- *spi*：安全策略索引（Security Policy Index）值，合法的十进制值为 256～4 294 967 295。
- *key-encryption-type*：密钥类型。0（默认）——未加密的密钥值；7——已加密的密钥值。
- *key*：计算 MD5 值使用的密钥值。32 位十六进制数。

例如，在区域 1 上启用认证：

```
router(config-rtr)#area 1 authentication ipsec spi 1000 md5 12345678901234567890123456789012
```

提示：12.4(4)T 及以后的 IOS 支持 IPv6 认证。

5.5.6　配置优雅重启特性

优雅重启（Graceful Restart）特性可以避免当路由器重启动时，邻居路由器清除连路状态数据库，造成大量路由临时删除的现象。当路由器重启完成后，又需要同步数据库，产生大量的 LSA 洪泛。该特性可以增强网络的稳定性。

优雅重启特性为路由器设置了两种身份，一种身份是具备优雅重启特性的路由器（Graceful Restart Capable Router）；另一种身份是可以识别邻居优雅重启特性的路由器（Graceful Restart Aware Router）。

在路由器上开启优雅重启特性的命令是：

```
Router(config-rtr)# graceful-restart
```

在路由器上开启识别优雅重启特性的命令是：

```
Router(config-rtr)# graceful-restart helper
```

5.6　本 章 小 结

OSPFv3 是支持 IPv6 的链路状态型路由协议，因此它与支持 IPv4 的 OSPFv2 有许多相似之处。但 IPv6 与 IPv4 有本质的不同，所以，OSPFv3 与 OSPFv2 也存在许多不同的特性。本章介绍了这两种路由选择协议的相似和不同之外。从数据包角度来看，OSPFv3 使用具有独特结构的 Hello 包和 LSA。从 LSA 类型上来看，OSPFv3 重新命名了原来的 LSA，并增加了 LSA 类型 8 和类型 9。

本章还介绍了如何在路由器上实现 OSPFv3 的命令和步骤，以及检验 OSPFv3 运行状态的常用命令，并举例介绍了如何在以太网环境和 NBMA 环境中配置 OSPFv3。本章最后介绍了如何进行路由总结和启用接口认证和区域认证。

第 6 章　支持 IPv6 的 IS-IS 协议

支持 IPv6 的集成 IS-IS 路由协议属于内部网关路由协议（IGP），使用连路状态信息通告过程建立网络拓扑信息库，然后使用最短路径优先（SPF）算法计算路由表。支持 IPv6 的集成 IS-IS 路由协议利用单一路由算法支持多个网络地址族——包括 IPv6、IPv4 和 OSI。

IS-ISv6 扩展了 IS-ISv4 的功能，在其 LSP 中可以携带 IPv6 前缀，其他方面的功能和特点与在 IPv4 中相同。相应地，IOS 也扩展了命令，在 CLI 中可以配置 IPv6 地址等与 IPv6 相关的参数。

6.1　为支持 IPv6 扩展的 TLV

为支持 IPv6，需要定义新的 TLV。

6.1.1　IPv6 可达性 TLV

在 RFC1195 中为 IPv4 定义了 2 个可达性 TLV，一个是内部可达性 TLV；另一个是外部可达性 TLV。在 IPv6 中，定义了和 IPv4 网络中相同功能的 TLV——IPv6 可达性 TLV，并在该 TLV 中使用一个“外部”标志比特来表示该路由是来自外部的路由，比如重分布的路由。

IPv6 可达性 TLV 结构如图 6-1 所示。

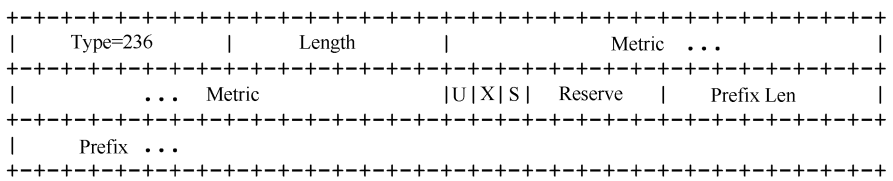

图 6-1　IPv6 可达性 TLV 结构

各字段含义如下所述。

- TLV 类型（Type）：236，一字节长；
- TLV 长度（Length）：一字节长；
- 度量值（Metric）：4 字节长；如果度量值大于定义的最大值（0xFE000000），SPF 在计算路由时，则不考虑该路由。
- U 比特：Up/Down 比特。当路由前缀从 L2 通告到 L1 时，该比特设置为 1，表示是从上层（Up Level）到下层（Down Level）的路由前缀；另外，该比特为 1 的路由只能往下层通告。路由器在选择优选路由时，也必须考虑该比特。它们的优先排序是：L1 Up > L2 Up > L2 Down > L Down。如果优先级相同，则比较度量值。如果不支持等值路径负载分担，则任选一个最佳路由。

- X 比特：外部起源比特。如果路由前缀是从其他路由协议重分布进入 IS-IS 的，该比特设置为 1；重分布到其他路由协议的 IS-IS 路由前缀也把该比特设置为 1。
- S 比特：携带子 TLV 标志。如果该比特为 0，表示后面没有携带子 TLV；如果为 1，表示后面携带子 TLV。子 TLV 的结构和父 TLV 的结构相同。
- 保留位（Reserve）：该 5 比特保留。
- 前缀长度（Prefix Len）：前缀长度。前缀占用的字节数使用公式为前缀字节数=（前缀长度值+7）/ 8 计算，然后取整数。
- 前缀（Prefix）：IPv6 路由前缀。

6.1.2　IPv6 接口地址 TLV

IPv6 接口地址 TLV 结构如图 6-2 所示：

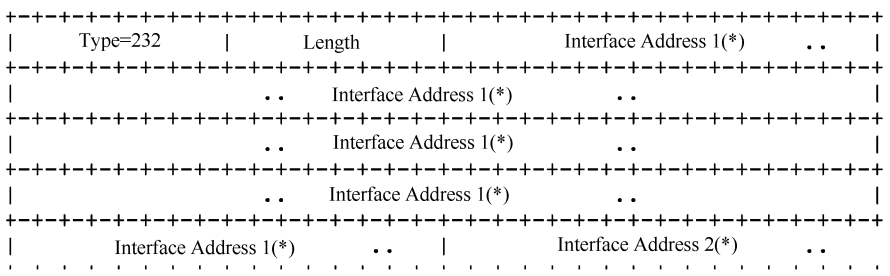

图 6-2　IPv6 接口地址 TLV 结构

- TLV 类型（Type）：232，一字节长。
- 长度（Length）：一字节。
- 接口地址（Interface Address）：接口 IPv6 地址；对于 Hello PDU，接口地址使用本链路地址；对于 LSP，接口地址使用非本链路地址。

6.1.3　网络层协议 ID TLV

支持 IPv6 的 IS-IS 必须在通告的 NLPID（Network Layer Protocol ID）TLV 中增加 IPv6 值，支持 IPv6 的 NLPID 值为 142。（报文结构参见《路由技术（IPv4 版）》）

6.1.4　多拓扑 TLV

在一个 IS-IS 路由域内运行的多个独立的 IP 拓扑称为多拓扑 IS-IS。接口上形成的邻接关系分别属于不同的拓扑，因此，需要在 IIH 中增加新的 TLV 宣告接口所属的拓扑。标准的 IP 拓扑定义为 MT#0，如果接口仅运行在单一的 IP 拓扑中，该值也可以不宣告，接收端默认为 MT#0。

多拓扑 TLV（Multi-Topology TLV）类型代码为 229，在 IIH 和 LSP 片段 0 中通告。MT TLV 最多可以宣告 127 个 MT。不携带 MT TLV 的 Hello 和 LSP 片段 0，被认为是 MT#0。如果路由器通告 MT TLV，它必须通告所有的 MT TLV，包括 MT #0。

MT TLV 的结构如图 6-3 所示。

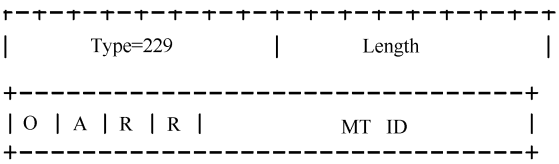

图 6-3　MT TLV

- 类型=229，1 字节。
- 长度=数值字段的长度值，1 字节。
- 数值部分 2 字节，包括，
  - O 比特：表示过载，仅在非 MT#0 的 LSP 片段 0 中有效，其他情况该比特应该设置为 0，并在接收端忽略；
  - A 比特：表示 ATTACH，仅在非 MT#0 的 LSP 片段 0 中有效，其他情况该比特应该设置为 0，并在接收端忽略；
  - R 比特：保留比特，值为 0，忽略；
  - MT ID：12 比特，MT ID 号。

6.1.5　MT IS TLV

该 TLV 类型代码为 222，是在 TLV-22（扩展 IS 可达性 TLV）的前面增加 2 字节的 MT ID 构成的，结构如图 6-4 所示。

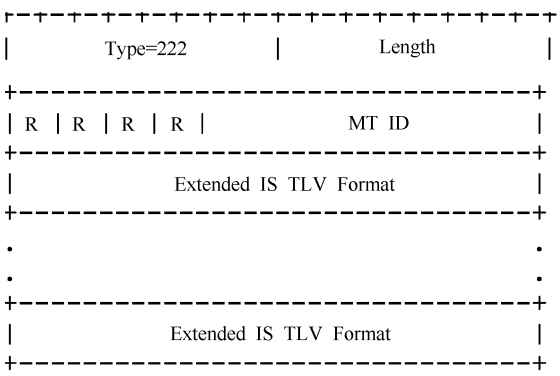

图 6-4　MT IS TLV 结构

如果 MT ID=0，则忽略该 TLV。扩展 IS TLV 的格式和类型 22 格式相同。

该类型 TLV 可以重复多次。

6.1.6　MT IPv4 可达前缀 TLV

该 TLV 类型代码为 235，是在 TLV-135 的前面增加 2 字节的 MT ID 构成，结构如图 6-5 所示。

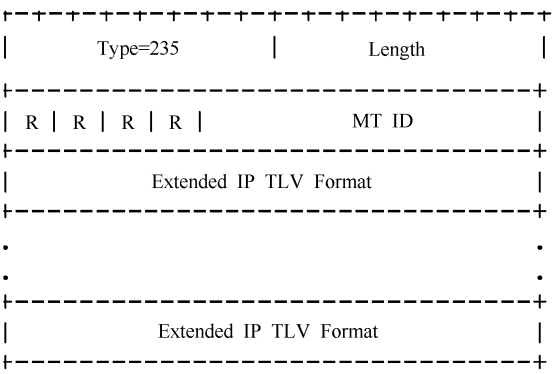

图 6-5　TLV 235 结构

该 TLV 可以重复多次。

6.1.7　MT IPv6 可达前缀 TLV

该 TLV 类型代码为 237，是在 TLV-236 的前面增加 2 字节的 MT ID 构成，结构如图 6-6 所示。

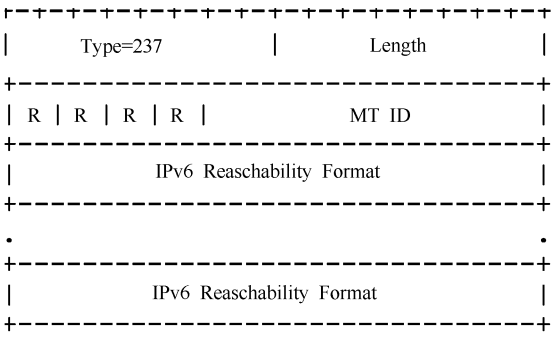

图 6-6　TLV 237 结构

该 TLV 可以重复多次。

6.2　单拓扑模式和多拓扑模式

单拓扑模式（Single-Topology Mode）在允许配置 IPv6 协议的同时，也可以为 IPv4 和 CLNS 提供路由功能。

单拓扑模式要求在同一个区域内的路由器必须支持相同的网络层协议，或者都仅支持 IPv6，或者都仅支持 IPv4，或者都同时支持双协议。

在接口上启用的 IS-IS 路由进程也必须是一致的，或者单 IPv6 协议进程，或者单 IPv4 协议进程，或者双协议进程，或者三协议进程，即 IPv6、IPv4 和 OSI。

在单拓扑模式下，如果接口配置了 IPv6 和 IPv4 地址，并且都运行在了 IS-IS 进程中，那么，它们运行的 Level 也必须相同，即或者都在 L1，或者都在 L2。

在上述限制下，运行在单拓扑模式下的路由器，其接口地址可能的情况是：所有接口都仅配置 IPv6 地址；或者所有接口都仅配置 IPv4 地址；或者所有接口都配置两种地址。

在单拓扑模式下，只有一个 SPF，同时为 IPv6、IPv4 和 OSI 计算路由，即 IPv6 IS-IS 和 IPv4 IS-IS 共用一个拓扑。

单拓扑支持新旧两种类型的 TLV，IPv6 和 IPv4 使用相同的度量，且 IPv6 仅支持宽度量，因此 IPv4 也需要使用宽度量。

IS-ISv6 多拓扑模式（Multi-Topology Mode）允许 IS-IS 在区域内或域内维护多个独立的拓扑，比如 IPv4 拓扑和 IPv6 拓扑。每个拓扑运行一个 SPF，为自己的拓扑提供路由和网络连通性保障。这种特性打破了单拓扑模式的一些限制，路由器接口配置的地址集不必一致，每个接口支持的网络层协议也可以是不同的。

在多拓扑模式下，可以单独为 IPv4 和 IPv6 定义度量值。

多拓扑模式下必须使用宽度量，因为 IPv6 仅支持宽度量。

在区域内或域内的路由器必须使用相同的拓扑模式，要么单模式，要么多模式。运行在多模式下的路由器不识别运行单模式的路由器具有支持 IPv6 的属性，因此会导致 IPv6 的路由黑洞。

为了灵活地从单模式迁移到多模式，IOS 提供了一个中间模式，称为转变模式（Transition Mode）。当从单拓扑模式升级到多拓扑模式时，路由器继续运行在单拓扑模式下，LSP 携带单拓扑和多拓扑两种 TLV。当所有路由器都升级到多拓扑模式后，可以停止转变模式，单拓扑模式也不在生效，路由器运行在多拓扑模式下。

提示：IOS 版本 12.2(15)T 以后支持多拓扑模式。

6.3　配置单拓扑单协议 IS-ISv6

本例使用的拓扑如图 6-7 所示。

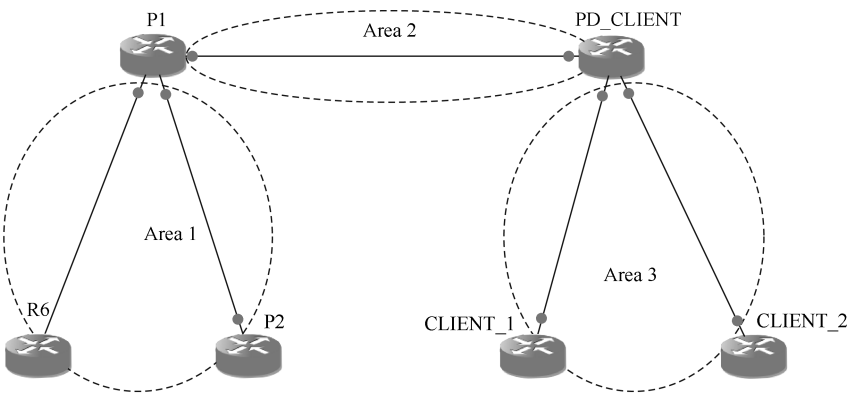

图 6-7　配置单拓扑单协议的拓扑

在此拓扑所示的环境中，每台路由器只启用 IPv6 路由协议。

6.3.1　基本配置

（1）路由器 R6 的相关配置

```
interface Loopback1
 no ip address
 ipv6 address 2013:1010:0:3::/64 eui-64
 ipv6 router isis a1
!
interface Loopback2
 no ip address
 ipv6 address 2013:1010:0:4::/64 eui-64
 ipv6 router isis a1
!
interface GigabitEthernet0/0
 no ip address
 duplex full
 speed 1000
 media-type gbic
 negotiation auto
 ipv6 address 2013:715:1624::/64 eui-64
 ipv6 router isis a1
!
router isis a1
 net 49.0001.0000.0000.0006.00
 metric-style wide
```

在 R6 上设置了 2 个 Loopback 接口，为的是产生更多的 IPv6 路由条目。

在 g0/0 接口上配置的 IPv6 地址与 P1 互联接口的地址并不在同一个网段上。这样，在 P1 的路由表里应该有这条路由（是 L1 路由）。

路由器 R6 设计在区域 1 内，启用了一个 IS-IS 进程。由于 IPv6 IS-IS 仅支持宽度量类型，所以使用了命令 metric-style wide。

默认地，进程是 L1/2 属性。

查看邻居关系，显示如下：

```
R6#sh clns neighbors

System Id      Interface   SNPA            State  Holdtime  Type Protocol
P1             Gi0/0       ca04.3f79.0008  Up     28        L1   IS-IS
```

（2）路由器 P2 的相关配置

```
interface Loopback1
 no ip address
 ipv6 address 2006:1010:0:5::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
interface Loopback2
 no ip address
 ipv6 address 2006:1010:0:6::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 address 2013:127:1647:4::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
router isis a1
 net 49.0001.0000.0000.0001.00
 metric-style wide
```

路由器 P2 和路由器 R6 同处于一个区域，它们的配置相同。

查看邻居关系，显示如下：

```
P2#sh clns neighbors

System Id      Interface   SNPA            State  Holdtime  Type  Protocol
P1             Fa0/0       ca04.3f79.001d  Up     8         L1    IS-IS
```

（3）路由器 P1 的相关配置

```
interface Loopback1
 no ip address
 ipv6 address 2013:1010:0:1::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
interface Loopback2
 no ip address
```

```
 ipv6 address 2013:1010:0:2::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
interface GigabitEthernet0/0
 no ip address
 duplex full
 speed 1000
 media-type gbic
 negotiation auto
 ipv6 address 2013:715:16::/64 eui-64
 ipv6 router isis a1
!
interface FastEthernet1/0
 no ip address
 duplex auto
 speed auto
 ipv6 address 2010:AB8:0:1::1/64
 ipv6 enable
 ipv6 router isis a2
!
interface FastEthernet1/1
 no ip address
 duplex auto
 speed auto
 ipv6 address 2013:127:1647:4::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
router isis a2
 net 49.0002.0000.0000.0002.00
 is-type level-2-only
 metric-style wide
!
router isis a1
 net 49.0001.0000.0000.0002.00
 is-type level-1
 metric-style wide
```

根据设计，路由器 P1 的 f1/0 接口运行在区域 2 中，接口 g0/0 和 f1/1 运行在区域 1 中。因此，采用了在路由器上启用 2 个进程，这 2 个进程的区域地址不同。

由于区域间路由是 L2，因此，区域 2 进程被优化为 Level-2-only 进程。

查看邻居关系，显示如下：

```
P1#sh clns neighbors
Area a2:
System Id      Interface   SNPA            State  Holdtime  Type  Protocol
PD_client      Fa1/0       cc03.3f73.0000  Up     7         L2    IS-IS
Area a1:
System Id      Interface   SNPA            State  Holdtime  Type  Protocol
P2             Fa1/1       ca02.3f72.0008  Up     22        L1    IS-IS
R6             Gi0/0       ca05.3f88.0008  Up     8         L1    IS-IS
```

（4）路由器 PD_Client 的相关配置

```
interface Loopback1
 no ip address
 ipv6 address 2013:1010:0:7::/64 eui-64
 ipv6 router isis a2
!
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 address autoconfig default
 ipv6 enable
 ipv6 router isis a2
 isis circuit-type level-2-only
!
interface FastEthernet1/0
 no ip address
 duplex auto
 speed auto
 ipv6 enable
 ipv6 router isis a3
!
interface FastEthernet3/0
 no ip address
 duplex auto
 speed auto
 ipv6 enable
 ipv6 router isis a3
!
router isis a2
```

```
 net 49.0002.0000.0000.0003.00
 metric-style wide
!
router isis a3
 net 49.0003.0000.0000.0003.00
 is-type level-1
 metric-style wide
```

在设计的拓扑中，该路由器和 P1 的地位是相同的，所以也配置了两个 IS-IS 进程，并分别在不同的区域里。

f0/0 使用自动地址配置，并优化链路类型为 Level-2-only。

因为 IS-IS 路由协议依靠本链路地址建立邻接关系、交换连路状态信息，所以该路由器其他的物理接口上都没有配置 IPv6 地址。

查看邻居关系，显示如下：

```
Area a2:
System Id      Interface   SNPA            State  Holdtime  Type  Protocol
P1             Fa0/0       ca04.3f79.001c  Up     23        L2    IS-IS

Area a3:
System Id      Interface   SNPA            State  Holdtime  Type  Protocol
client_2       Fa3/0       cc01.3f71.0000  Up     28        L1    IS-IS
client_1       Fa1/0       cc00.3f71.0000  Up     23        L1    IS-IS
```

（5）路由器 Client_1 的相关配置

```
interface Loopback1
 no ip address
 ipv6 address 2013:1010:0:8::/64 eui-64
 ipv6 router isis a3
!
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 enable
 ipv6 router isis a3
!
router isis a3
 net 49.0003.0000.0000.0008.00
 metric-style wide
```

物理接口没有配置 IPv6 地址。

默认地，IS-IS 进程是 L1/2 进程，也没有优化。

查看邻居关系如下：

```
client_1#sh clns neighbors

System Id      Interface   SNPA            State  Holdtime  Type  Protocol
PD_client      Fa0/0       cc03.3f73.0010  Up     7         L1    IS-IS
client_1#
```

（6）路由器 Client_2 的相关配置

```
interface Loopback1
 no ip address
 ipv6 address 2013:1010:0:9::/64 eui-64
 ipv6 router isis a3
!
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 enable
 ipv6 router isis a3
!
router isis a3
 net 49.0003.0000.0000.0009.00
 is-type level-1
 metric-style wide
```

把 IS-IS 进程优化为 Level-1 进程。

查看邻居关系，显示如下：

```
client_2#sh clns neighbors

System Id      Interface   SNPA            State  Holdtime  Type  Protocol
PD_client      Fa0/0       cc03.3f73.0030  Up     9         L1    IS-IS
client_2#
```

6.3.2　检查运行状态

1．查看 IS-IS 邻居关系建立情况

下面显示的是在 P1 上看到的邻居：

```
P1#sh isis neighbors

Area a2:
System Id      Type  Interface  IP Address      State Holdtime  Circuit Id
```

```
PD_client        L2    Fa1/0                    UP     8          PD_client.02
Area a1:
System Id        Type  Interface  IP Address    State Holdtime    Circuit Id
P2               L1    Fa1/1                    UP     23         P1.04
R6               L1    Gi0/0                    UP     9          R6.03
P1#
```

由于没有配置 IPv4 地址，所以 IP Address 项无内容显示。

下面显示的是在 PD_client 上看到的邻居：

```
PD_Client#sh isis neighbors

Area a2:
System Id        Type  Interface  IP Address    State Holdtime    Circuit Id
P1               L2    Fa0/0                    UP     27         PD_Client.02

Area a3:
System Id        Type  Interface  IP Address    State Holdtime    Circuit Id
client_2         L1    Fa3/0                    UP     25         PD_Client.02
client_1         L1    Fa1/0                    UP     25         PD_Client.01
PD_Client#
```

上述命令可以使用 detail 选项，显示更详细的内容。例如：

```
PD_Client#sh isis neighbors detail

Area a2:
System Id        Type Interface IP Address      State Holdtime    Circuit Id
P1               L2    Fa0/0                    UP     26         PD_Client.02
  Area Address(es): 49.0002
  SNPA: ca04.3f79.001c
  IPv6 Address(es): FE80::C804:3FFF:FE79:1C
  State Changed: 00:54:52
  LAN Priority: 64
  Format: Phase V

Area a3:
System Id        Type Interface IP Address      State Holdtime    Circuit Id
client_2         L1    Fa3/0                    UP     23         PD_Client.02
  Area Address(es): 49.0003
  SNPA: cc01.3f71.0000
  IPv6 Address(es): FE80::CE01:3FFF:FE71:0
  State Changed: 02:43:42
  LAN Priority: 64
  Format: Phase V
```

```
client_1          L1    Fa1/0                    UP     24          PD_Client.01
  Area Address(es): 49.0003
  SNPA: cc00.3f71.0000
  IPv6 Address(es): FE80::CE00:3FFF:FE71:0
  State Changed: 02:51:29
  LAN Priority: 64
  Format: Phase V
PD_Client#
```

2．查看 DIS 的选举情况

以太链路对于 IS-IS 来说是广播型链路，在广播型链路上需要选举 DIS。下面显示的是 PD_Client 这台路由器的 3 个接口上 DIS 选举的情况。

```
PD_Client#sh clns interface
FastEthernet0/0 is up, line protocol is up
  Checksums enabled, MTU 1497, Encapsulation SAP
  ERPDUs enabled, min. interval 10 msec.
  CLNS fast switching enabled
  CLNS SSE switching disabled
  DEC compatibility mode OFF for this interface
  Next ESH/ISH in 21 seconds
  Routing Protocol: IS-IS (a2)
    Circuit Type: level-2
    Interface number 0x1, local circuit ID 0x2
    Level-2 Metric: 10, Priority: 64, Circuit ID: PD_Client.02
    DR ID: PD_Client.02
    Level-2 IPv6 Metric: 10
    Number of active level-2 adjacencies: 1
    Next IS-IS LAN Level-2 Hello in 1 seconds
FastEthernet1/0 is up, line protocol is up
  Checksums enabled, MTU 1497, Encapsulation SAP
  ERPDUs enabled, min. interval 10 msec.
  CLNS fast switching enabled
  CLNS SSE switching disabled
  DEC compatibility mode OFF for this interface
  Next ESH/ISH in 34 seconds
  Routing Protocol: IS-IS (a3)
    Circuit Type: level-1-2
    Interface number 0x0, local circuit ID 0x1
    Level-1 Metric: 10, Priority: 64, Circuit ID: PD_Client.01
    DR ID: PD_Client.01
    Level-1 IPv6 Metric: 10
```

```
    Number of active level-1 adjacencies: 1
    Next IS-IS LAN Level-1 Hello in 977 milliseconds
FastEthernet3/0 is up, line protocol is up
  Checksums enabled, MTU 1497, Encapsulation SAP
  ERPDUs enabled, min. interval 10 msec.
  CLNS fast switching enabled
  CLNS SSE switching disabled
  DEC compatibility mode OFF for this interface
  Next ESH/ISH in 9 seconds
  Routing Protocol: IS-IS (a3)
    Circuit Type: level-1-2
    Interface number 0x1, local circuit ID 0x2
    Level-1 Metric: 10, Priority: 64, Circuit ID: PD_Client.02
    DR ID: PD_Client.02
    Level-1 IPv6 Metric: 10
    Number of active level-1 adjacencies: 1
    Next IS-IS LAN Level-1 Hello in 1 seconds
PD_Client#
```

3．查看接口的链路类型

例如：

```
P1# sh clns int f1/0
FastEthernet1/0 is up, line protocol is up
  Checksums enabled, MTU 1497, Encapsulation SAP
  ERPDUs enabled, min. interval 10 msec.
  CLNS fast switching enabled
  CLNS SSE switching disabled
  DEC compatibility mode OFF for this interface
  Next ESH/ISH in 7 seconds
  Routing Protocol: IS-IS (a2)
    Circuit Type: level-1-2
    Interface number 0x0, local circuit ID 0x1
    Level-2 Metric: 10, Priority: 64, Circuit ID: PD_Client.02
    DR ID: PD_Client.02
    Level-2 IPv6 Metric: 10
    Number of active level-2 adjacencies: 1
    Next IS-IS LAN Level-2 Hello in 3 seconds
```

这里显示，P1 的 f1/0 接口（和 PD_Client 互连）的链路类型是 Level-1-2 的。但是，该接口运行的进程是 Level-2-only 进程，所以和邻居建立的是 L2 关系。

再看 PD_Client 的接口链路类型：

```
PD_Client#sh clns int f0/0
FastEthernet0/0 is up, line protocol is up
  Checksums enabled, MTU 1497, Encapsulation SAP
  ERPDUs enabled, min. interval 10 msec.
  CLNS fast switching enabled
  CLNS SSE switching disabled
  DEC compatibility mode OFF for this interface
  Next ESH/ISH in 21 seconds
  Routing Protocol: IS-IS (a2)
    Circuit Type: level-2
    Interface number 0x1, local circuit ID 0x2
    Level-2 Metric: 10, Priority: 64, Circuit ID: PD_Client.02
    DR ID: PD_Client.02
    Level-2 IPv6 Metric: 10
    Number of active level-2 adjacencies: 1
    Next IS-IS LAN Level-2 Hello in 850 milliseconds
PD_Client#
```

该接口的链路类型为 Level-2。但是，该接口运行的进程是 L1/2 进程。因此，也能与邻居建立 L2 关系。

偶然地，这种互连的接口链路类型不同，导致报文中携带的链路类型值不一致，因此控制台报告错误提示。例如，PD_Client 上有如下提示：

```
*Mar  1 02:23:25.471: %CLNS-3-BADPACKET: ISIS: LAN L2 hello, bad circuit type 0 from ca04.3f79.001c (FastEthernet0/0)
```

将 P1 的 f1/0 接口的链路类型修改为 Level-2 即可消除此提示。

4．查看路由表

（1）R6 的路由表

```
R6#sh ipv6 route isis
IPv6 Routing Table - 14 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       U - Per-user Static route, M - MIPv6
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
       D - EIGRP, EX - EIGRP external
I1  ::/0 [115/10]
      via FE80::C804:3FFF:FE79:8, GigabitEthernet0/0
I1  2006:1010:0:5::/64 [115/30]
      via FE80::C804:3FFF:FE79:8, GigabitEthernet0/0
I1  2006:1010:0:6::/64 [115/30]
```

```
        via FE80::C804:3FFF:FE79:8, GigabitEthernet0/0
I1   2013:127:1647:4::/64 [115/20]
        via FE80::C804:3FFF:FE79:8, GigabitEthernet0/0
I1   2013:715:16::/64 [115/20]
        via FE80::C804:3FFF:FE79:8, GigabitEthernet0/0
I1   2013:1010:0:1::/64 [115/20]
        via FE80::C804:3FFF:FE79:8, GigabitEthernet0/0
I1   2013:1010:0:2::/64 [115/20]
        via FE80::C804:3FFF:FE79:8, GigabitEthernet0/0
R6#
```

（2）P2 的路由表

```
P2#sh ipv6 route isis
IPv6 Routing Table - 14 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       U - Per-user Static route, M - MIPv6
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
       D - EIGRP, EX - EIGRP external
I1   ::/0 [115/10]
        via FE80::C804:5DFF:FEAC:1D, FastEthernet0/0
I1   2013:715:16::/64 [115/20]
        via FE80::C804:5DFF:FEAC:1D, FastEthernet0/0
I1   2013:715:1624::/64 [115/30]
        via FE80::C804:5DFF:FEAC:1D, FastEthernet0/0
I1   2013:1010:0:1::/64 [115/20]
        via FE80::C804:5DFF:FEAC:1D, FastEthernet0/0
I1   2013:1010:0:2::/64 [115/20]
        via FE80::C804:5DFF:FEAC:1D, FastEthernet0/0
I1   2013:1010:0:3::/64 [115/30]
        via FE80::C804:5DFF:FEAC:1D, FastEthernet0/0
I1   2013:1010:0:4::/64 [115/30]
        via FE80::C804:5DFF:FEAC:1D, FastEthernet0/0
P2#
```

（3）P1 的路由表

```
P1#sh ipv6 route isis
IPv6 Routing Table - 19 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       U - Per-user Static route, M - MIPv6
```

```
        I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
        O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
        ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
        D - EIGRP, EX - EIGRP external
I1  2006:1010:0:5::/64 [115/20]
      via FE80::C802:3FFF:FE72:8, FastEthernet1/1
I1  2006:1010:0:6::/64 [115/20]
      via FE80::C802:3FFF:FE72:8, FastEthernet1/1
I1  2013:715:1624::/64 [115/20]
      via FE80::C805:3FFF:FE88:8, GigabitEthernet0/0
I1  2013:1010:0:3::/64 [115/20]
      via FE80::C805:3FFF:FE88:8, GigabitEthernet0/0
I1  2013:1010:0:4::/64 [115/20]
      via FE80::C805:3FFF:FE88:8, GigabitEthernet0/0
I2  2013:1010:0:7::/64 [115/20]
      via FE80::CE03:3FFF:FE73:0, FastEthernet1/0
I2  2013:1010:0:8::/64 [115/30]
      via FE80::CE03:3FFF:FE73:0, FastEthernet1/0
I2  2013:1010:0:9::/64 [115/30]
      via FE80::CE03:3FFF:FE73:0, FastEthernet1/0
P1#
```

（4）PD_Client 的路由表

```
PD_Client#sh ipv6 route isis
IPv6 Routing Table - 14 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
        U - Per-user Static route
        I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
        O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
        ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
I2  2006:1010:0:5::/64 [115/30]
      via FE80::C804:3FFF:FE79:1C, FastEthernet0/0
I2  2006:1010:0:6::/64 [115/30]
      via FE80::C804:3FFF:FE79:1C, FastEthernet0/0
I2  2013:715:1624::/64 [115/30]
      via FE80::C804:3FFF:FE79:1C, FastEthernet0/0
I2  2013:1010:0:3::/64 [115/30]
      via FE80::C804:3FFF:FE79:1C, FastEthernet0/0
I2  2013:1010:0:4::/64 [115/30]
      via FE80::C804:3FFF:FE79:1C, FastEthernet0/0
I1  2013:1010:0:8::/64 [115/20]
```

```
        via FE80::CE00:3FFF:FE71:0, FastEthernet1/0
I1    2013:1010:0:9::/64 [115/20]
        via FE80::CE01:3FFF:FE71:0, FastEthernet3/0
PD_Client#
```

（5）Client_2 的路由表

```
client_2#sh ipv6 route isis
IPv6 Routing Table - 6 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
          U - Per-user Static route
          I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
          O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
          ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
I1    ::/0 [115/10]
        via FE80::CE03:3FFF:FE73:30, FastEthernet0/0
I1    2013:1010:0:8::/64 [115/30]
        via FE80::CE03:3FFF:FE73:30, FastEthernet0/0
client_2#
```

（6）Client_1 的路由表

```
client_1#sh ipv6 route isis
IPv6 Routing Table - 6 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
          U - Per-user Static route
          I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
          O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
          ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
I1    ::/0 [115/10]
        via FE80::CE03:5DFF:FE98:10, FastEthernet0/0
I1    2013:1010:0:9::/64 [115/30]
        via FE80::CE03:5DFF:FE98:10, FastEthernet0/0
client_1#
```

从路由表看出，所有区域内的路由器都自动产生一条默认路由，下一跳指向区域边界路由器。

5．查看链路状态数据库

（1）PD_Client 路由器的数据库

```
PD_Client#sh isis database

Area a2:
IS-IS Level-1 Link State Database:
```

```
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime      ATT/P/OL
PD_Client.00-00     * 0x00000003   0xD513        1093              1/0/0
IS-IS Level-2 Link State Database:
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime      ATT/P/OL
P1.00-00              0x00000005   0x839D        999               0/0/0
PD_Client.00-00     * 0x00000008   0x02B0        973               0/0/0
PD_Client.02-00     * 0x00000003   0x883B        1022              0/0/0

Area a3:
IS-IS Level-1 Link State Database:
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime      ATT/P/OL
PD_Client.00-00     * 0x00000005   0x8ACB        968               1/0/0
PD_Client.01-00     * 0x00000003   0xEF46        968               0/0/0
PD_Client.02-00     * 0x00000003   0xFC37        1125              0/0/0
client_1.00-00      0x00000005     0x49B9        966               0/0/0
client_2.00-00      0x00000004     0x0BF2        802               0/0/0
PD_Client#
```

因为区域 a2 的进程是 L1/2 进程，所以有 L1 和 L2 链路状态库。区域 a3 修改成了 L1 进程，所以只有 L1 链路状态库。

（2）Client_1 的链路状态库

```
client_1#sh isis database

IS-IS Level-1 Link State Database:
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime      ATT/P/OL
PD_Client.00-00       0x00000005   0x8ACB        680               1/0/0
PD_Client.01-00       0x00000003   0xEF46        680               0/0/0
PD_Client.02-00       0x00000003   0xFC37        837               0/0/0
client_1.00-00      * 0x00000005   0x49B9        682               0/0/0
client_2.00-00        0x00000004   0x0BF2        514               0/0/0
IS-IS Level-2 Link State Database:
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime      ATT/P/OL
client_1.00-00      * 0x00000001   0x5219        687               0/0/0
client_1#
```

该路由器上的 IS-IS 进程是 L1/2 进程（默认状态），所以有 L2 数据库。但是，它只和邻居建立了 L1 邻接关系，没有 L2 邻接关系，所以路由表中只有 L1 路由。

下面查看一下 L2 数据库中的内容：

```
client_1#sh isis database l2 detail

IS-IS Level-2 Link State Database:
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime      ATT/P/OL
client_1.00-00      * 0x00000001   0x5219        657               0/0/0
```

```
    Area Address: 49.0003
    NLPID:          0x8E
    Hostname: client_1
    IPv6 Address: 2013:1010:0:8:CE00:5DFF:FE96:0
    Metric: 10          IPv6 2013:1010:0:8::/64
    Metric: 30          IPv6 2013:1010:0:9::/64
  client_1#
```

从 L2 链路状态数据库中的内容看，client_1 把自己的 L1 路由信息放入了 L2 数据库，这说明了为什么跨区域的路由是以 L2 的路由属性传递的原因。

其他路由器的链路状态数据库这里不再展示了。

6.4　配置单拓扑多协议 IS-ISv6

6.4.1　理解单拓扑模式特性

6.2 节介绍了单拓扑模式的特性，比如，在接口上启用的 IS-IS 路由进程也必须是一致的：或者是单 IPv6 协议进程，或者是单 IPv4 协议进程，或者是双协议进程，或者是三协议进程，即 IPv6、IPv4 和 OSI。

下面演示的是该特性在路由器上的表现。

以路由器 P1 为例。

只把 P1 的 Loopback1 接口上配置一个 IP 地址，并把该接口运行在 IS-IS 进程 a1 中，其他配置不需要更改，然后观察路由器的反应。

P1 的 Loopback1 接口配置后的情况如下：

```
  interface Loopback1
   ip address 1.1.1.1 255.255.255.0
   ip router isis a1
   ipv6 address 2013:1010:0:1::/64 eui-64
   ipv6 enable
   ipv6 router isis a1
  !
```

查看 P1 的邻居关系：

```
  P1#sh clns neighbors
  Area a2:
  System Id      Interface   SNPA            State  Holdtime  Type  Protocol
  PD_Client      Fa1/0       cc03.5d98.0000  Up     8         L2    IS-IS
  Area a1:
  System Id      Interface   SNPA            State  Holdtime  Type  Protocol
  R6             Gi0/0       ca05.5dad.0008  Up     251       IS    ES-IS
  P2             Fa1/1       ca02.5d97.0008  Up     250       IS    ES-IS
  P1#
```

输出显示，P1 和 R6、P2 的 CLNS 邻居关系变换成了 ES-IS 关系，而不是之前的 IS-IS 关系。

再看 IS-IS 邻居情况：

```
P1#sh isis neighbors

Area a2:
System Id      Type Interface IP Address      State Holdtime    Circuit Id
PD_Client      L2   Fa1/0                     UP    8           PD_Client.02
Area a1:
System Id      Type Interface IP Address      State Holdtime Circuit Id
P1#
```

输出显示，在区域 a1 中没有 IS-IS 邻居关系。

再观察 R6 和 P2：

```
R6#sh clns neighbors

System Id      Interface   SNPA            State  Holdtime  Type Protocol
P1             Gi0/0       ca04.5dac.0008  Init   8         L1   IS-IS
R6#

P2#sh clns neighbors

System Id      Interface   SNPA            State  Holdtime  Type Protocol
P1             Fa0/0       ca04.5dac.001d  Init   8         L1   IS-IS
P2#
```

它们和 P1 的邻居关系没有建立起来，停留在了 Init 状态。

既然邻接关系都没有建立起来，肯定也不会有路由信息的交换，也就没有 IPv6 路由了。

6.4.2　配置单拓扑多协议 IS-ISv6

配置单拓扑多协议，需要满足单拓扑的限制，即在每个接口上都配置 IPv4 和 IPv6 的地址，并运行在相应的 IS-IS 进程中。下面以图 6-8 为例介绍单拓扑多协议的配置。

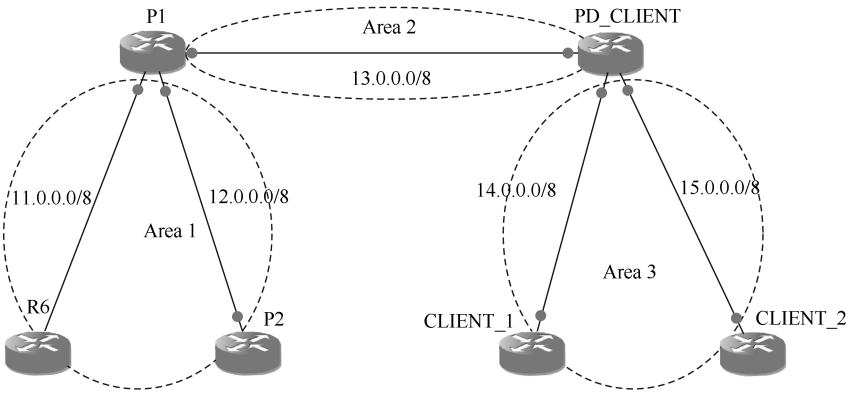

图 6-8　单拓扑多协议拓扑

在上节配置的基础上，每台路由器都配置 IPv4 地址，并运行在相应 IS-IS 进程中。

（1）路由器 P1 的配置

```
interface Loopback1
 ip address 1.1.1.1 255.255.255.0
 ip router isis a1
 ipv6 address 2013:1010:0:1::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
interface Loopback2
 ip address 2.2.2.2 255.255.255.255
 ip router isis a1
 ipv6 address 2013:1010:0:2::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
interface GigabitEthernet0/0
 ip address 11.0.0.1 255.0.0.0
 ip router isis a1
 duplex full
 speed 1000
 media-type gbic
 negotiation auto
 ipv6 address 2013:715:16::/64 eui-64
 ipv6 router isis a1
 clns router isis a1
!
interface FastEthernet1/0
 ip address 13.0.0.1 255.0.0.0
 ip router isis a2
 duplex auto
 speed auto
 ipv6 address 2010:AB8:0:1::1/64
 ipv6 enable
 ipv6 router isis a2
 clns router isis a2
!
interface FastEthernet1/1
 ip address 12.0.0.1 255.0.0.0
 ip router isis a1
```

```
 duplex auto
 speed auto
 ipv6 address 2013:127:1647:4::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
router isis a2
 net 49.0002.0000.0000.0002.00
 is-type level-2-only
 metric-style wide
!
router isis a1
 net 49.0001.0000.0000.0002.00
 is-type level-1
 metric-style wide
```

（2）路由器 R6 的配置

```
interface Loopback1
 ip address 3.3.3.3 255.255.255.255
 ip router isis a1
 ipv6 address 2013:1010:0:3::/64 eui-64
 ipv6 router isis a1
!
interface Loopback2
 ip address 4.4.4.4 255.255.255.255
 ip router isis a1
 ipv6 address 2013:1010:0:4::/64 eui-64
 ipv6 router isis a1
!
interface GigabitEthernet0/0
 ip address 11.0.0.6 255.0.0.0
 ip router isis a1
 duplex full
 speed 1000
 media-type gbic
 negotiation auto
 ipv6 address 2013:715:1624::/64 eui-64
 ipv6 router isis a1
!
router isis a1
 net 49.0001.0000.0000.0006.00
 metric-style wide
```

（3）路由器 P2 的配置

```
interface Loopback1
 ip address 5.5.5.5 255.255.255.255
 ip router isis a1
 ipv6 address 2006:1010:0:5::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
interface Loopback2
 ip address 6.6.6.6 255.255.255.255
 ip router isis a1
 ipv6 address 2006:1010:0:6::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
interface FastEthernet0/0
 ip address 12.0.0.2 255.0.0.0
 ip router isis a1
 duplex auto
 speed auto
 ipv6 address 2013:127:1647:4::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
 clns router isis a1
!
router isis a1
 net 49.0001.0000.0000.0001.00
 metric-style wide
```

（4）路由器 PD_Client 的配置

```
interface Loopback1
 ip address 7.7.7.7 255.255.255.255
 ip router isis a2
 ipv6 address 2013:1010:0:7::/64 eui-64
 ipv6 router isis a2
!
interface FastEthernet0/0
 ip address 13.0.0.3 255.0.0.0
 ip router isis a2
 duplex auto
 speed auto
```

```
 ipv6 address autoconfig default
 ipv6 enable
 ipv6 router isis a2
 isis circuit-type level-2-only
!
interface FastEthernet1/0
 ip address 14.0.0.3 255.0.0.0
 ip router isis a3
 duplex auto
 speed auto
 ipv6 enable
 ipv6 router isis a3
!
interface FastEthernet3/0
 ip address 15.0.0.3 255.0.0.0
 ip router isis a3
 duplex auto
 speed auto
 ipv6 enable
 ipv6 router isis a3
!
router isis a2
 net 49.0002.0000.0000.0003.00
 metric-style wide
!
router isis a3
 net 49.0003.0000.0000.0003.00
 is-type level-1
 metric-style wide
```

（5）路由器 Client_1 的配置

```
interface Loopback1
 ip address 8.8.8.8 255.255.255.255
 ip router isis a3
 ipv6 address 2013:1010:0:8::/64 eui-64
 ipv6 router isis a3
!
interface FastEthernet0/0
 ip address 14.0.0.1 255.0.0.0
 ip router isis a3
 duplex auto
```

```
 speed auto
 ipv6 enable
 ipv6 router isis a3
!
router isis a3
 net 49.0003.0000.0000.0008.00
 metric-style wide
```

（6）路由器 Client_2 的配置

```
interface Loopback1
 ip address 9.9.9.9 255.255.255.255
 ip router isis a3
 ipv6 address 2013:1010:0:9::/64 eui-64
 ipv6 router isis a3
!
interface FastEthernet0/0
 ip address 15.0.0.2 255.0.0.0
 ip router isis a3
 duplex auto
 speed auto
 ipv6 enable
 ipv6 router isis a3
!
router isis a3
 net 49.0003.0000.0000.0009.00
 is-type level-1
 metric-style wide
```

6.4.3　检查运行状态

1．查看邻居关系

（1）P1 上的邻居关系

```
P1#sh clns neighbors
Area a2:
System Id      Interface   SNPA            State  Holdtime  Type  Protocol
PD_Client      Fa1/0       cc03.5d98.0000  Up     7         L2    IS-IS
Area a1:
System Id      Interface   SNPA            State  Holdtime  Type  Protocol
R6             Gi0/0       ca05.5dad.0008  Up     7         L1    IS-IS
P2             Fa1/1       ca02.5d97.0008  Up     29        L1    IS-IS
```

```
P1#

P1#sh isis neighbors

Area a2:
System Id      Type Interface   IP Address      State Holdtime   Circuit Id
PD_Client      L2   Fa1/0       13.0.0.3        UP    9          PD_Client.02
Area a1:
System Id      Type Interface   IP Address      State Holdtime   Circuit Id
R6             L1   Gi0/0       11.0.0.6        UP    9          R6.03
P2             L1   Fa1/1       12.0.0.2        UP    29         P1.04
P1#
```

（2）PD_Client 上的邻居关系

```
PD_Client#sh clns neighbors

Area a2:
System Id      Interface   SNPA            State  Holdtime  Type  Protocol
P1             Fa0/0       ca04.5dac.001c  Up     26        L2    IS-IS

Area a3:
System Id      Interface   SNPA            State  Holdtime  Type  Protocol
client_2       Fa3/0       cc01.5d96.0000  Up     25        L1    IS-IS
client_1       Fa1/0       cc00.5d96.0000  Up     25        L1    IS-IS
PD_Client#
```

```
PD_Client#sh isis neighbors

Area a2:
System Id      Type Interface   IP Address      State Holdtime   Circuit Id
P1             L2   Fa0/0       13.0.0.1        UP    26         PD_Client.02

Area a3:
System Id      Type Interface   IP Address      State Holdtime   Circuit Id
client_2       L1   Fa3/0       15.0.0.2        UP    27         PD_Client.02
client_1       L1   Fa1/0       14.0.0.1        UP    26         PD_Client.01
PD_Client#
```

2．查看链路状态数据库

由于数据库输出的内容比较多，这里只展示 PD_Client 的数据库：

```
PD_Client#sh isis database    detail

Area a2:
IS-IS Level-1 Link State Database:
LSPID                 LSP Seq Num   LSP Checksum   LSP Holdtime      ATT/P/OL
PD_Client.00-00     * 0x0000001A    0xA8DF         746               1/0/0
  Area Address: 49.0002
  NLPID:          0xCC 0x8E
  Hostname: PD_Client
  IP Address:     7.7.7.7
  Metric: 10      IP 7.7.7.7/32
  IPv6 Address: 2013:1010:0:7:CE03:5DFF:FE98:0
  Metric: 10      IPv6 2013:1010:0:7::/64
IS-IS Level-2 Link State Database:
LSPID                 LSP Seq Num   LSP Checksum   LSP Holdtime      ATT/P/OL
P1.00-00              0x0000002E    0x1E59         1089              0/0/0
  Area Address: 49.0001
  Area Address: 49.0002
  NLPID:          0x81 0xCC 0x8E
  Hostname: P1
  IP Address:     13.0.0.1
  IPv6 Address: 2010:AB8:0:1::1
  Metric: 10      IS-Extended PD_Client.02
  Metric: 10      IP 1.1.1.0/24
  Metric: 10      IP 2.2.2.2/32
  Metric: 20      IP 3.3.3.3/32
  Metric: 20      IP 4.4.4.4/32
  Metric: 20      IP 5.5.5.5/32
  Metric: 20      IP 6.6.6.6/32
  Metric: 10      IP 11.0.0.0/8
  Metric: 10      IP 12.0.0.0/8
  Metric: 10      IP 13.0.0.0/8
  Metric: 20      IPv6 2006:1010:0:5::/64
  Metric: 20      IPv6 2006:1010:0:6::/64
  Metric: 10      IPv6 2010:AB8:0:1::/64
  Metric: 20      IPv6 2013:715:1624::/64
  Metric: 20      IPv6 2013:1010:0:3::/64
  Metric: 20      IPv6 2013:1010:0:4::/64
PD_Client.00-00     * 0x00000027    0x1B76         470               0/0/0
  Area Address: 49.0003
  Area Address: 49.0002
```

```
  NLPID:          0xCC 0x8E
  Hostname: PD_Client
  IP Address:     7.7.7.7
  Metric: 10      IP 13.0.0.0/8
  IPv6 Address: 2013:1010:0:7:CE03:5DFF:FE98:0
  Metric: 10      IPv6 2010:AB8:0:1::/64
  Metric: 10      IS-Extended PD_Client.02
  Metric: 10      IP 7.7.7.7/32
  Metric: 20      IP 8.8.8.8/32
  Metric: 20      IP 9.9.9.9/32
  Metric: 10      IP 14.0.0.0/8
  Metric: 10      IP 15.0.0.0/8
  Metric: 10      IPv6 2013:1010:0:7::/64
  Metric: 20      IPv6 2013:1010:0:8::/64
  Metric: 20      IPv6 2013:1010:0:9::/64
PD_Client.02-00   * 0x00000019   0x5C51   735   0/0/0
  Metric: 0       IS-Extended PD_Client.00
  Metric: 0       IS-Extended P1.00

Area a3:
IS-IS Level-1 Link State Database:
LSPID             LSP Seq Num   LSP Checksum   LSP Holdtime   ATT/P/OL
PD_Client.00-00   * 0x00000021   0x6993   644   1/0/0
  Area Address: 49.0003
  NLPID:          0xCC 0x8E
  Hostname: PD_Client
  IP Address:     15.0.0.3
  Metric: 10      IP 14.0.0.0/8
  Metric: 10      IP 15.0.0.0/8
  Metric: 10      IS-Extended PD_Client.02
  Metric: 10      IS-Extended PD_Client.01
PD_Client.01-00   * 0x00000019   0xC35C   532   0/0/0
  Metric: 0       IS-Extended PD_Client.00
  Metric: 0       IS-Extended client_1.00
PD_Client.02-00   * 0x00000019   0xD04D   1163   0/0/0
  Metric: 0       IS-Extended PD_Client.00
  Metric: 0       IS-Extended client_2.00
client_1.00-00    0x0000001F   0x84EA   601   0/0/0
  Area Address: 49.0003
  NLPID:          0xCC 0x8E
  Hostname: client_1
```

```
    IP Address:     8.8.8.8
    Metric: 10              IP 14.0.0.0/8
    Metric: 10              IP 8.8.8.8/32
    IPv6 Address: 2013:1010:0:8:CE00:5DFF:FE96:0
    Metric: 10              IPv6 2013:1010:0:8::/64
    Metric: 10              IS-Extended PD_Client.01
client_2.00-00          0x0000001E      0xD989          1041                0/0/0
    Area Address: 49.0003
    NLPID:          0xCC 0x8E
    Hostname: client_2
    IP Address:     9.9.9.9
    Metric: 10              IP 15.0.0.0/8
    Metric: 10              IP 9.9.9.9/32
    IPv6 Address: 2013:1010:0:9:CE01:5DFF:FE96:0
    Metric: 10              IPv6 2013:1010:0:9::/64
    Metric: 10              IS-Extended PD_Client.02
PD_Client#
```

关于数据库的详细解释，请参阅《路由技术（IPv4 版）》

3．查看路由表

（1）P1 上的路由表

```
P1#sh ip route isis
      3.0.0.0/32 is subnetted, 1 subnets
i L1     3.3.3.3 [115/20] via 11.0.0.6, GigabitEthernet0/0
      4.0.0.0/32 is subnetted, 1 subnets
i L1     4.4.4.4 [115/20] via 11.0.0.6, GigabitEthernet0/0
      5.0.0.0/32 is subnetted, 1 subnets
i L1     5.5.5.5 [115/20] via 12.0.0.2, FastEthernet1/1
      6.0.0.0/32 is subnetted, 1 subnets
i L1     6.6.6.6 [115/20] via 12.0.0.2, FastEthernet1/1
      7.0.0.0/32 is subnetted, 1 subnets
i L2     7.7.7.7 [115/20] via 13.0.0.3, FastEthernet1/0
      8.0.0.0/32 is subnetted, 1 subnets
i L2     8.8.8.8 [115/30] via 13.0.0.3, FastEthernet1/0
      9.0.0.0/32 is subnetted, 1 subnets
i L2     9.9.9.9 [115/30] via 13.0.0.3, FastEthernet1/0
i L2 14.0.0.0/8 [115/20] via 13.0.0.3, FastEthernet1/0
i L2 15.0.0.0/8 [115/20] via 13.0.0.3, FastEthernet1/0
P1#
```

```
P1#
P1#sh ipv6 route isis
IPv6 Routing Table - 19 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       U - Per-user Static route, M - MIPv6
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
       D - EIGRP, EX - EIGRP external
I1   2006:1010:0:5::/64 [115/20]
       via FE80::C802:5DFF:FE97:8, FastEthernet1/1
I1   2006:1010:0:6::/64 [115/20]
       via FE80::C802:5DFF:FE97:8, FastEthernet1/1
I1   2013:715:1624::/64 [115/20]
       via FE80::C805:5DFF:FEAD:8, GigabitEthernet0/0
I1   2013:1010:0:3::/64 [115/20]
       via FE80::C805:5DFF:FEAD:8, GigabitEthernet0/0
I1   2013:1010:0:4::/64 [115/20]
       via FE80::C805:5DFF:FEAD:8, GigabitEthernet0/0
I2   2013:1010:0:7::/64 [115/20]
       via FE80::CE03:5DFF:FE98:0, FastEthernet1/0
I2   2013:1010:0:8::/64 [115/30]
       via FE80::CE03:5DFF:FE98:0, FastEthernet1/0
I2   2013:1010:0:9::/64 [115/30]
       via FE80::CE03:5DFF:FE98:0, FastEthernet1/0
P1#
```

注意：IPv6 和 IPv4 的路由度量值是相同的，这是单拓扑模式的特性。

（2）R6 上的路由表

```
R6#sh ip route isis
      1.0.0.0/24 is subnetted, 1 subnets
i L1     1.1.1.0 [115/20] via 11.0.0.1, GigabitEthernet0/0
      2.0.0.0/32 is subnetted, 1 subnets
i L1     2.2.2.2 [115/20] via 11.0.0.1, GigabitEthernet0/0
      5.0.0.0/32 is subnetted, 1 subnets
i L1     5.5.5.5 [115/30] via 11.0.0.1, GigabitEthernet0/0
      6.0.0.0/32 is subnetted, 1 subnets
i L1     6.6.6.6 [115/30] via 11.0.0.1, GigabitEthernet0/0
i L1 12.0.0.0/8 [115/20] via 11.0.0.1, GigabitEthernet0/0
i*L1 0.0.0.0/0 [115/10] via 11.0.0.1, GigabitEthernet0/0
```

```
R6#
R6#
R6#sh ipv6 route isis
IPv6 Routing Table - 14 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       U - Per-user Static route, M - MIPv6
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
       D - EIGRP, EX - EIGRP external
I1   ::/0 [115/10]
       via FE80::C804:5DFF:FEAC:8, GigabitEthernet0/0
I1   2006:1010:0:5::/64 [115/30]
       via FE80::C804:5DFF:FEAC:8, GigabitEthernet0/0
I1   2006:1010:0:6::/64 [115/30]
       via FE80::C804:5DFF:FEAC:8, GigabitEthernet0/0
I1   2013:127:1647:4::/64 [115/20]
       via FE80::C804:5DFF:FEAC:8, GigabitEthernet0/0
I1   2013:715:16::/64 [115/20]
       via FE80::C804:5DFF:FEAC:8, GigabitEthernet0/0
I1   2013:1010:0:1::/64 [115/20]
       via FE80::C804:5DFF:FEAC:8, GigabitEthernet0/0
I1   2013:1010:0:2::/64 [115/20]
       via FE80::C804:5DFF:FEAC:8, GigabitEthernet0/0
R6#
```

为节约篇幅，其他路由器的路由表不一一列出了。

4．测试连通性

在 R6 上测试到达 Client_2 的 Loopback1 接口（2013:1010:0:9:CE01:5DFF:FE96:0）连通性：

```
R6#ping ipv6 2013:1010:0:9:CE01:5DFF:FE96:0

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 2013:1010:0:9:CE01:5DFF:FE96:0, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 36/54/68 ms
R6#
R6#
R6#ping ip 9.9.9.9

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 9.9.9.9, timeout is 2 seconds:
```

```
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 28/59/88 ms
R6#
```

6.5　配置多拓扑 IS-ISv6

6.5.1　拓扑设计

为演示如何配置多拓扑 IS-ISv6 而设计的拓扑如图 6-9 所示。

在区域 1 中，路由器 R6 只运行 IPv4 拓扑中；路由器 P2 只运行 IPv6 拓扑中。

路由器 P1 的 2 个 Loopback 接口既运行在 IPv4 拓扑中，也运行在 IPv6 拓扑中，它们设计在区域 2 中运行。

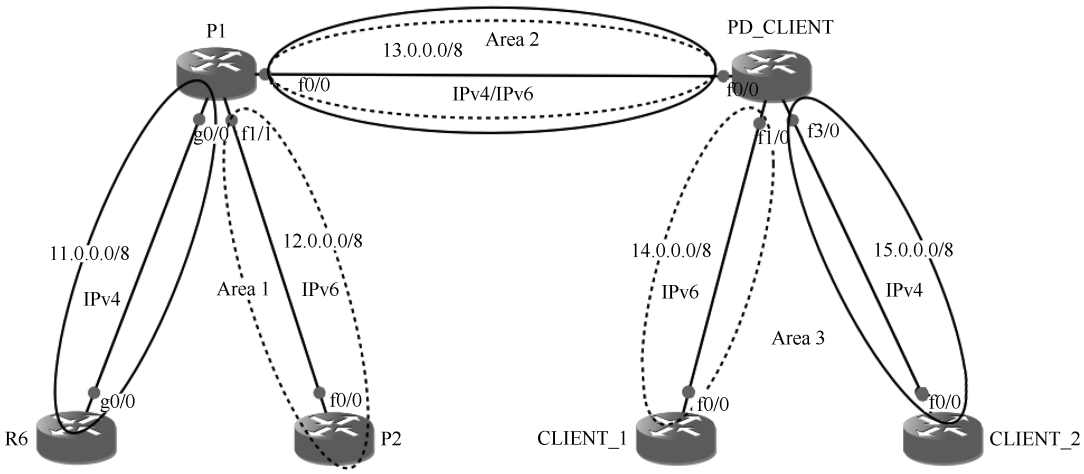

图 6-9　多拓扑演示拓扑

- 在区域 2 中的 2 台路由器物理接口运行在 IP4 和 IPv6 拓扑中；
- 路由器 PD-Client 的 Loopback1 接口运行 IPv6 拓扑中，设计运行在区域 2。
- 在区域 3 中，client_1 只运行在 IPv6 拓扑中；client_2 只运行在 IPv4 拓扑中。

6.5.2　迁移至多拓扑模式

从单拓扑模式迁移至多拓扑模式需要以下几个步骤。

第一步：在每台路由器上的每个 IS-IS 进程下，配置 IPv6 多拓扑模式迁移命令。

例如，配置 P1：

```
P1(config)#router isis a2
P1(config-router)#address-family ipv6 unicast
P1(config-router-af)#multi-topology transition
P1(config-router-af)#exit
P1(config-router)#exit
```

```
P1(config)#router isis a1
P1(config-router)#address-family ipv6 unicast
P1(config-router-af)#multi-topology transition
P1(config-router-af)#exit
```

当完成配置后，查看它们的邻居关系，例如，P1 当前的邻居关系如下：

```
P1#sh clns neighbors

Area a2:
System Id      Interface   SNPA            State  Holdtime  Type  Protocol
PD_Client      Fa1/0       cc03.0ca0.0000  Up     8         L2    M-ISIS

Area a1:
System Id      Interface   SNPA            State  Holdtime  Type  Protocol
R6             Gi0/0       ca05.0cb5.0008  Up     9         L1    M-ISIS
P2             Fa1/1       ca02.0c9f.0008  Up     27        L1    M-ISIS
P1#
```

输出显示，已经建立起多拓扑邻接关系（M-ISIS）。

第二步：根据设计，修改接口配置。

修改完成后各路由器的配置如下：

（1）R6 的配置

```
interface Loopback1
 ip address 3.3.3.3 255.255.255.255
 ip router isis a1
 ipv6 address 2013:1010:0:3::/64 eui-64
!
interface Loopback2
 ip address 4.4.4.4 255.255.255.255
 ip router isis a1
 ipv6 address 2013:1010:0:4::/64 eui-64
!
interface GigabitEthernet0/0
 ip address 11.0.0.6 255.0.0.0
 ip router isis a1
 duplex full
 speed 1000
 media-type gbic
 negotiation auto
!
router isis a1
 net 49.0001.0000.0000.0006.00
```

```
 metric-style wide
 !
 address-family ipv6
  multi-topology transition
 exit-address-family
```

（2）P2 的配置

```
interface Loopback1
 ip address 5.5.5.5 255.255.255.255
 ipv6 address 2006:1010:0:5::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
interface Loopback2
 ip address 6.6.6.6 255.255.255.255
 ipv6 address 2006:1010:0:6::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
interface FastEthernet0/0
 ip address 12.0.0.2 255.0.0.0
 duplex auto
 speed auto
 ipv6 address 2013:127:1647:4::/64 eui-64
 ipv6 enable
 ipv6 router isis a1
!
interface FastEthernet0/1
 ip address 16.0.0.1 255.0.0.0
 duplex auto
 speed auto
 ipv6 address 2013:1:1:1::/64 eui-64
 ipv6 router isis a1
 no keepalive
!
router isis a1
 net 49.0001.0000.0000.0001.00
 metric-style wide
 !
 address-family ipv6
 multi-topology transition
 exit-address-family
```

（3）P1 的配置

```
interface Loopback1
 ip address 1.1.1.1 255.255.255.0
 ip router isis a2
 ipv6 address 2013:1010:0:1::/64 eui-64
 ipv6 enable
 ipv6 router isis a2
!
interface Loopback2
 ip address 2.2.2.2 255.255.255.255
 ip router isis a2
 ipv6 address 2013:1010:0:2::/64 eui-64
 ipv6 enable
 ipv6 router isis a2
!
interface GigabitEthernet0/0
 ip address 11.0.0.1 255.0.0.0
 ip router isis a1
 duplex full
 speed 1000
 media-type gbic
 negotiation auto
 ipv6 address 2013:715:16::/64 eui-64
!
interface FastEthernet1/0
 ip address 13.0.0.1 255.0.0.0
 ip router isis a2
 duplex auto
 speed auto
 ipv6 address 2010:AB8:0:1::1/64
 ipv6 enable
 ipv6 router isis a2
 clns router isis a2
 isis circuit-type level-2-only
!
interface FastEthernet1/1
 ip address 12.0.0.1 255.0.0.0
 duplex auto
 speed auto
 ipv6 address 2013:127:1647:4::/64 eui-64
 ipv6 enable
```

```
 ipv6 router isis a1
!
router isis a2
 net 49.0002.0000.0000.0002.00
 metric-style wide
 !
 address-family ipv6
 multi-topology transition
 exit-address-family
!
router isis a1
 net 49.0001.0000.0000.0002.00
 is-type level-1
 metric-style wide
 !
 address-family ipv6
 multi-topology transition
 exit-address-family
```

（4）PD_Client 的配置

```
interface Loopback1
 ip address 7.7.7.7 255.255.255.255
 ip router isis a2
 ipv6 address 2013:1010:0:7::/64 eui-64
 ipv6 router isis a2
!
interface FastEthernet0/0
 ip address 13.0.0.7 255.0.0.0
 ip router isis a2
 duplex auto
 speed auto
 ipv6 address autoconfig default
 ipv6 enable
 ipv6 router isis a2
 isis circuit-type level-2-only
!
interface FastEthernet1/0
 ip address 14.0.0.7 255.0.0.0
 duplex auto
 speed auto
 ipv6 enable
```

```
 ipv6 router isis a3
!
interface FastEthernet3/0
 ip address 15.0.0.7 255.0.0.0
 ip router isis a3
 duplex auto
 speed auto
 ipv6 address provider ::1/64
 ipv6 enable
!
router isis a2
 net 49.0002.0000.0000.0003.00
 metric-style wide
 !
 address-family ipv6
 multi-topology transition
 exit-address-family
!
router isis a3
 net 49.0003.0000.0000.0003.00
 is-type level-1
 metric-style wide
 !
 address-family ipv6
 multi-topology transition
 exit-address-family
```

（5）Client_1 的配置

```
interface Loopback1
 ip address 8.8.8.8 255.255.255.255
 ipv6 address 2013:1010:0:8::/64 eui-64
 ipv6 router isis a3
!
interface FastEthernet0/0
 ip address 14.0.0.1 255.0.0.0
 duplex auto
 speed auto
 ipv6 address autoconfig
 ipv6 enable
 ipv6 nd other-config-flag
 ipv6 router isis a3
```

```
!
router isis a3
 net 49.0003.0000.0000.0008.00
 metric-style wide
 !
 address-family ipv6
 multi-topology transition
 exit-address-family
```

（6）Client_2 的配置

```
interface Loopback1
 ip address 9.9.9.9 255.255.255.255
 ip router isis a3
 ipv6 address 2013:1010:0:9::/64 eui-64
!
interface FastEthernet0/0
 ip address 15.0.0.2 255.0.0.0
 ip router isis a3
 duplex auto
 speed auto
 ipv6 enable
!
router isis a3
 net 49.0003.0000.0000.0009.00
 is-type level-1
 metric-style wide
!
address-family ipv6
 multi-topology transition
 exit-address-family
```

第三步：等待建立完整路由表。

例如，查看 P 的路由表：

```
P1#sh ip route isis
      3.0.0.0/32 is subnetted, 1 subnets
i L1     3.3.3.3 [115/20] via 11.0.0.6, GigabitEthernet0/0
      4.0.0.0/32 is subnetted, 1 subnets
i L1     4.4.4.4 [115/20] via 11.0.0.6, GigabitEthernet0/0
      7.0.0.0/32 is subnetted, 1 subnets
i L2     7.7.7.7 [115/20] via 13.0.0.7, FastEthernet1/0
      9.0.0.0/32 is subnetted, 1 subnets
i L2     9.9.9.9 [115/30] via 13.0.0.7, FastEthernet1/0
```

```
i L2 14.0.0.0/8 [115/20] via 13.0.0.7, FastEthernet1/0
i L2 15.0.0.0/8 [115/20] via 13.0.0.7, FastEthernet1/0
P1#
P1#
P1#sh ipv6 route isis
IPv6 Routing Table - 16 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       U - Per-user Static route, M - MIPv6
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       O - OSPF intra, OI - OSPF inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
       D - EIGRP, EX - EIGRP external
I1   2006:1010:0:5::/64 [115/20]
       via FE80::C802:CFF:FE9F:8, FastEthernet1/1
I1   2006:1010:0:6::/64 [115/20]
       via FE80::C802:CFF:FE9F:8, FastEthernet1/1
I1   2013:1:1:1::/64 [115/20]
       via FE80::C802:CFF:FE9F:8, FastEthernet1/1
I2   2013:1010:0:7::/64 [115/20]
       via FE80::CE03:CFF:FEA0:0, FastEthernet1/0
I2   2013:1010:0:8::/64 [115/30]
       via FE80::CE03:CFF:FEA0:0, FastEthernet1/0
P1#
```

测试一下连通性。

从 R6 上测试到达 Client_2 的 loopback 接口的连通性:

```
R6#ping 9.9.9.9

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 9.9.9.9, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 40/47/56 ms
R6#
```

从 P2 上测试到达 Client_1 的 loopback 接口的连通性:

```
P2#ping ipv6 2013:1010:0:8:CE00:CFF:FE9E:0

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 2013:1010:0:8:CE00:CFF:FE9E:0, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 12/41/76 ms
P2#
```

第四步：输入 multi-topology，完成迁移。

完成后查看一下链路状态数据库，以 P1 为例：

```
P1#sh isis database detail

Area a2:
IS-IS Level-1 Link State Database:
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime      ATT/P/OL
P1.00-00            * 0x00000002   0xBA27        658               1/0/0
  Area Address: 49.0002
  Topology:       IPv4 (0x0) IPv6 (0x4002 ATT)
  NLPID:          0x81 0xCC 0x8E
  Hostname: P1
  IP Address:     2.2.2.2
  Metric: 10          IP 1.1.1.0/24
  Metric: 10          IP 2.2.2.2/32
  IPv6 Address: 2013:1010:0:2:C804:CFF:FEB4:6
  Metric: 10          IPv6 (MT-IPv6) 2013:1010:0:1::/64
  Metric: 10          IPv6 (MT-IPv6) 2013:1010:0:2::/64
  Metric: 0           ES P1
IS-IS Level-2 Link State Database:
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime      ATT/P/OL
P1.00-00            * 0x00000021   0x96D1        688               0/0/0
  Area Address: 49.0001
  Area Address: 49.0002
  Topology:       IPv4 (0x0) IPv6 (0x2)
  NLPID:          0x81 0xCC 0x8E
  Hostname: P1
  IP Address:     2.2.2.2
  Metric: 10          IP 13.0.0.0/8
  IPv6 Address: 2013:1010:0:2:C804:CFF:FEB4:6
  Metric: 10          IPv6 (MT-IPv6) 2010:AB8:0:1::/64
  Metric: 10          IS-Extended PD_Client.01
  Metric: 10          IS (MT-IPv6) PD_Client.01
  Metric: 10          IP 1.1.1.0/24
  Metric: 10          IP 2.2.2.2/32
  Metric: 20          IP 3.3.3.3/32
  Metric: 20          IP 4.4.4.4/32
  Metric: 10          IP 11.0.0.0/8
  Metric: 20          IPv6 (MT-IPv6) 2006:1010:0:5::/64
  Metric: 20          IPv6 (MT-IPv6) 2006:1010:0:6::/64
  Metric: 20          IPv6 (MT-IPv6) 2013:1:1:1::/64
```

```
  Metric: 10          IPv6 (MT-IPv6) 2013:1010:0:1::/64
  Metric: 10          IPv6 (MT-IPv6) 2013:1010:0:2::/64
PD_Client.00-00       0x00000022    0x31FF          762               0/0/0
  Area Address: 49.0003
  Area Address: 49.0002
  Topology:       IPv4 (0x0) IPv6 (0x2)
  NLPID:          0xCC 0x8E
  Hostname: PD_Client
  IP Address:     7.7.7.7
  Metric: 10          IP 13.0.0.0/8
  IPv6 Address: 2013:1010:0:7:CE03:CFF:FEA0:0
  Metric: 10          IPv6 (MT-IPv6) 2010:AB8:0:1::/64
  Metric: 10          IS-Extended PD_Client.01
  Metric: 10          IS (MT-IPv6) PD_Client.01
  Metric: 10          IP 7.7.7.7/32
  Metric: 20          IP 9.9.9.9/32
  Metric: 10          IP 15.0.0.0/8
  Metric: 10          IPv6 (MT-IPv6) 2013:1010:0:7::/64
  Metric: 20          IPv6 (MT-IPv6) 2013:1010:0:8::/64
PD_Client.01-00       0x0000000C    0x7D3E          835               0/0/0
  Metric: 0           IS-Extended PD_Client.00
  Metric: 0           IS-Extended P1.00

Area a1:
IS-IS Level-1 Link State Database:
LSPID                 LSP Seq Num   LSP Checksum  LSP Holdtime      ATT/P/OL
P2.00-00              0x00000014    0xDECF        925               0/0/0
  Area Address: 49.0001
  Topology:       IPv6 (0x2)
  NLPID:          0x81 0x8E
  Hostname: P2
  IPv6 Address: 2006:1010:0:6:C802:CFF:FE9F:8
  Metric: 10          IPv6 (MT-IPv6) 2013:127:1647:4::/64
  Metric: 10          IPv6 (MT-IPv6) 2013:1:1:1::/64
  Metric: 10          IPv6 (MT-IPv6) 2006:1010:0:5::/64
  Metric: 10          IPv6 (MT-IPv6) 2006:1010:0:6::/64
  Metric: 10          IS-Extended P1.02
  Metric: 10          IS (MT-IPv6) P1.02
  Metric: 0           ES P2
P1.00-00              * 0x00000018    0xB7E2        824               1/0/0
  Area Address: 49.0001
```

```
    Topology:        IPv4 (0x0) IPv6 (0x4002 ATT)
    NLPID:           0xCC 0x8E
    Hostname: P1
    IP Address:      11.0.0.1
    Metric: 10       IP 11.0.0.0/8
    IPv6 Address: 2013:127:1647:4:C804:CFF:FEB4:1D
    Metric: 10       IPv6 (MT-IPv6) 2013:127:1647:4::/64
    Metric: 10       IS-Extended R6.03
    Metric: 10       IS (MT-IPv6) P1.02
  P1.02-00           * 0x0000000F   0x43EE        686           0/0/0
    Metric: 0        IS-Extended P1.00
    Metric: 0        IS-Extended P2.00
  R6.00-00             0x00000011   0xF93C        1141          0/0/0
    Area Address: 49.0001
    NLPID:           0xCC
    Hostname: R6
    IP Address:      4.4.4.4
    Metric: 10       IP 11.0.0.0/8
    Metric: 10       IP 3.3.3.3/32
    Metric: 10       IP 4.4.4.4/32
    Metric: 10       IS-Extended R6.03
  R6.03-00             0x0000000D   0x58D1        904           0/0/0
    Metric: 0        IS-Extended R6.00
    Metric: 0        IS-Extended P1.00
  P1#
```

从输出的数据库看出，P1 和 PD_Client 都通告了 IPv4 和 MT-IPv6 路由，而 R6 和 P2 分别只通告了 IPv4 路由和 MT-IPv6 路由。

6.6 本 章 小 结

本章讨论了为支持 IPv6 而对 IS-IS 进行 TLV 扩展。这些扩展包括网络层协议扩展 TLV、接口地址 TLV、多拓扑网络层可达 TLV。有了这些 TLV 扩展，IS-IS 协议才可以携带 IPv6 相关的路由信息，才可以支持为不同路由协议运行一个独立的拓扑。

IS-ISv6 最大的特点是支持多种拓扑模式。

单拓扑模式要求在同一个区域内的路由器必须支持相同的网络层协议。在接口上启用的 IS-IS 路由进程也必须是一致的。它们运行的 Level 也必须相同，即或者都在 L1，或者都在 L2。

多拓扑模式允许 IS-IS 在区域内或域内维护多个独立的拓扑，比如 IPv4 拓扑和 IPv6 拓扑。每个拓扑运行一个 SPF，为自己的拓扑提供路由和网络连通性保障。这种特性打破了单拓扑模式的一些限制，路由器接口配置的地址集不必一致，每个接口支持的网络层协议也可以是不同的。

介绍了 IS-ISv6 原理之后，分别演示了如何配置单拓扑单协议、单拓扑多协议和如何迁移到多拓扑模式。

第 7 章　IPv6 BGP 序幕——BGPv4

本章讨论的边界网关协议（Border Gateway Protocol，BGP），其名字的由来与路由协议的分类有关。根据分类依据的不同，对路由选择协议的分类方法也有多种，其中一种分类方法的依据是自治系统（Autonomous System，AS）。自治系统的概念在 CCNA 的书籍中曾经介绍过，但是，当谈到 BGP 协议时需要对它做进一步的认识。

7.1　BGP 与 AS

根据路由协议是否运行在自治系统内，路由选择协议可以分为如下两种类型，如图 7-1 所示。

- 内部网关协议（Interior Gateway Protocol，IGP）——在自治系统内运行的路由选择协议称为内部网关协议。本章之前讨论的路由协议都属于 IGP。
- 外部网关协议（Exterior Gateway Protocol，EGP）——运行在自治系统之间的路由选择协议称为外部网关协议。边界网关协议就属于 EGP。

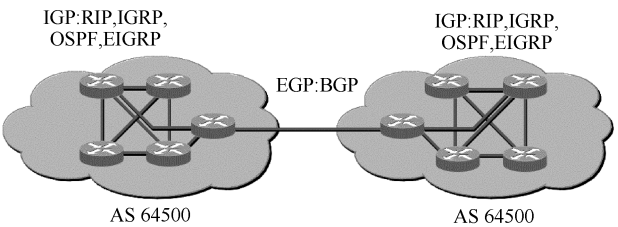

图 7-1　IGP 与 EGP

RFC1771 对自治系统的描述是这样的：自治系统的传统定义是指在单一技术管理下，采用一种内部网关协议和通用的度量在 AS 内转发数据包，并采用一种外部网关协议将数据包路由到其他 AS 的一组路由器。在这个传统的定义使用后，一个单一的自治系统内使用几个 IGP，而且，有时候使用几套度量已经普遍存在。这里使用“自治系统”这个术语强调了这样一个事实：尽管使用了多个 IGP 和多套度量，一个 AS 的管理展示给其他 AS 的是它有一个单一紧密的（Coherent）内部路由计划，并且，对那些经由该 AS 可达的目的地，AS 的管理还向其他 AS 呈现出一致画面。

BGP 用在自治系统之间，其主要功能是提供无环的域间（自治系统间）路由信息，并对进 / 出自治系统的数据依据特定的策略进行调控。

自治系统用一个 16 比特的数值标志，其十进制范围从 1～65 535。从 64 512～65 535 的编号是私用的，类似于 IP 地址中的私有地址。

另外一种协议分类方法是依据路由协议的运行原理，把路由协议分为距离矢量型路由协议、链路状态型路由协议和综合型路由协议。在本章以前谈到的路由协议中，OSPF、IS-IS 属于链路状态型路由协议，EIGRP 属于综合型路由协议，也称为高级距离矢量型路由协议。

BGP 也是一种带有许多增强特性的高级距离矢量型路由协议。大多数链路状态型路由协议都要求体系化设计网络，这主要是为了便于路由总结，而 BGP 则不要求体系化设计网络。

BGP 路由协议计算度量的方式与其他协议计算度量的方式有所不同。OSPF 使用链路开销（Cost）计算度量。IS-IS 除了可以使用链路开销（默认度量）计算度量外，还可以使用其他参数，如误差、延迟和花费计算度量。EIGRP 使用一种复合（带宽、延迟）的度量值计算公式。总之，这些路由协议最终计算出的度量值是一个正的自然数，而 BGP 使用的度量并非一个具体的数值。BGP 使用一系列属性衡量路径的优劣（远近），通过比较路由信息中携带的一些属性，最终展现出的是到达目的地所经由的所有 AS 的列表，因此也称为路径矢量。

提示：BGP 属性在 7.6 节讨论。

7.2　使用 BGP 时的考虑

自治系统之间的数据交换并非一定使用 BGP 路由协议提供路径信息，当 BGP 的影响被完全了解并且至少存在下列情况之一时，使用 BGP 才是合适的。

- AS 允许数据包穿过它到达其他自治系统。
- AS 有到达其他自治系统的多条连接。
- 必须对进入和离开 AS 的数据进行控制。
- 必须将 AS 与其 ISP 的数据区分开。
- 本地策略（Policy）与 ISP 策略不同。

BGP 被用在 ISP 之间进行通信和交换数据包，这些 ISP 间存在多条链路，并且有交换路由更新的策略（规则），BGP 在两个或多个自治系统间实施这些策略的协议。如果没有对 BGP 进行正确的控制和过滤，那么它可能潜在地会允许一个外部 AS 影响它的路由决策。

对于互连的自治系统来说，BGP 并不总是恰当的解决方案。例如，自治系统之间只存在一条路径，在这种情况下，默认路由或静态路由是很合适的。如果一定要使用 BGP 的话，只能是浪费路由器的 CPU 和内存资源。

如果存在下面情况中的一种或多种，不要使用 BGP。

- AS 内实施的路由策略与 ISP 的 AS 中实施的策略一致。
- 只有单一连接到另一 AS 或 Internet。
- 无须考虑路由策略或路由选择。
- 路由器缺乏 BGP 更新的内存或处理能力。
- 对路由过滤和 BGP 路径选择过程不十分了解。
- 自治系统之间的带宽较低。

在这些情况下，应使用静态路由。

7.3　BGP 概述

和 TCP/IP 协议堆栈中的其他路由协议一样，BGP 的路由信息由 IP 数据包携带。BGP 的独特特征是它在第四层使用 TCP 作为传输层协议，端口号是 179。这样可以提供面向连接的可

靠传输，并且因此不再附加任何其他的可靠机制。携带 BGP 信息的数据结构如图 7-2 所示。

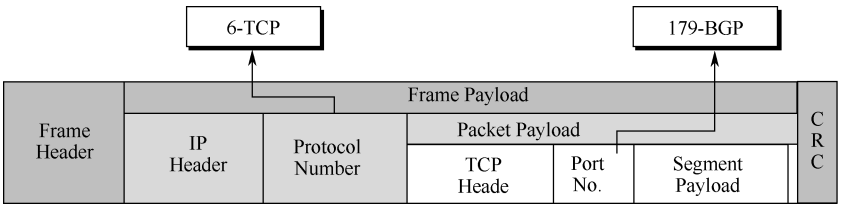

图 7-2　携带 BGP 信息的帧结构

运行 BGP 协议的路由器在交换路由信息之前必须建立起邻居关系，BGP 路由器可以通过建立一条 TCP 连接而成为邻居。已经建立起 TCP 连接的 BGP 路由器互相称对方为自己的对等体（Peers）或邻居。连接 BGP 对等体的链路可以在同一 AS 内，也可以在不同的 AS 内。

在同一个 AS 内的 BGP 对等体称为内部 BGP 对等体（Interior BGP，IBGP）。IBGP 对等体在 AS 内交换 BGP 信息。IBGP 邻居之间不必有直连的链路，只要它们相互间可达（能建立起 TCP 会话）即可。不同 AS 内的 BGP 对等体，称为外部 BGP 对等体（Exterior BGP，EBGP）。EBGP 在不同自治系统间交换 BGP 信息。EBGP 邻居之间一般要有直连链路。图 7-3 描述了 IBGP 关系和 EBGP 关系。

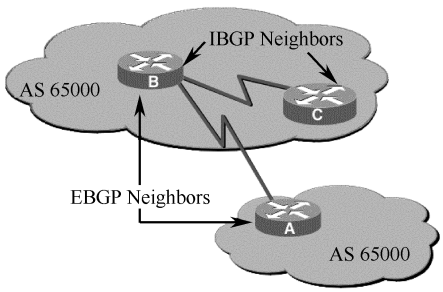

图 7-3　IBGP 和 EBGP

当邻接关系建立起来之后，邻接路由器之间将交换 BGP 路由信息，并把 BGP 路由信息保存在路由信息数据库（Routing Information Bases，RIB）中。该数据库与路由器中的 IGP 数据库是分开的，但它们之间可以在管理员的干预下（通过配置）互相交换路由信息。在第一次 BGP 数据库同步之后，BGP 路由器只需发送增量更新信息，所以，BGP 采用了触发更新机制。

BGP 用来描述网络可达性的信息被称为路径矢量，它由若干路径属性组成，包括路由到达目标网络经由的全部 AS 列表（路径列表）。该 AS 列表在构建无环路径中起着十分重要的作用，因为，运行 BGP 的路由器将不接受在 AS 列表中包含自己 AS 编号的路由更新信息。含有自己 AS 编号意味着该路径已经经过了本 AS，如果再接受它将会导致路由环路。

BGP 允许为数据怎样通过 AS 而定义策略（Policy）。路由器根据设置的路由策略或规则运行被称为基于策略的路由，设置的策略基于路由信息中承载的以及配置在路由器上的 BGP 属性工作。

BGP 规定，BGP 路由器只能够向相邻自治系统的对等体通告那些自己使用的路由。这条规则也反映了当前 Internet 采用“逐跳（Hop-by-hop）”的路由模式。BGP 能够支持任何与这种模式相符的策略。因为当前 Internet 仅使用“逐跳”的路由模式，所以，在当前的 Internet 上，BGP 是一种非常适用于 AS 间路由的路由选择协议。

7.4　BGP 消息类型

BGP 消息有固定的 19 字节报文头，如图 7-4 所示。

- Marker：接收者可以根据此推测一些信息，占 16 字节。

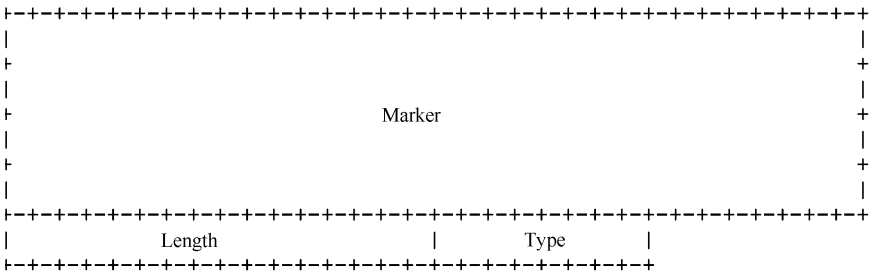

图 7-4　BGP 消息报文头

- Length：报文总长，占 2 字节。
- Type：报文类型，占 1 字节。

BGP 运行时将会使用以下消息类型：

- 打开（Open）消息，类型=1。
- 更新（Update）消息，类型=2。
- 通知（Notification）消息，类型=3。
- 存活（Keepalive）消息，类型=4。

（1）打开消息

当建立起一条 TCP 连接后，双方发送的第一条消息是打开消息。如果 Open 消息是可以接收的，路由器回送 Keepalive 消息对它进行确认。一旦 Open 消息被确认，双方可以交换 Update、Keepalive 和 Notification 消息。

OPEN 消息格式如图 7-5 所示。

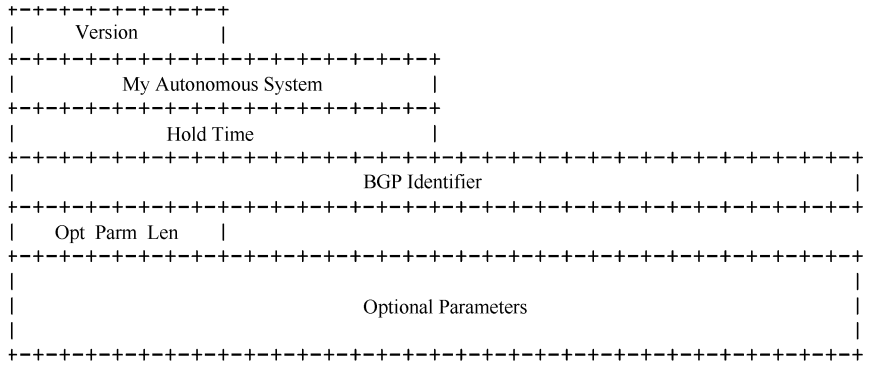

图 7-5　OPEN 报文结构

除了固定的 19 字节报文头外，打开消息包含以下内容：

- 版本（8 bit）——表示 BGP 协议版本号，当前版本是 4。
- AS 号（16 bit）——发送该消息的路由器所在的自治系统编号。
- 保持时间（16 bit）——连续 Keepalive 或更新消息之间最长等待的时间。路由器将从它配置的保持时间值和打开消息中的保持时间值中选择较小值使用，默认值为 180 秒。如果双方协商的保持时间为 0，那么将不发送 Keepalive 消息。
- BGP ID（32 bit）——即路由器 ID，代表发送方的 BGP 标志符。BGP 路由器 ID 的选

择方法与 OSPF 路由器 ID 的选择方法一样。

- 任选参数域长度（8bit）——表示任选参数域的总长，以字节为单位。
- 任选参数（变长）——可能包含的任选参数。

（2）更新消息

当 BGP 连接建立起来后，BGP 邻居就可以进行首次 BGP 路由信息的完全交换，当 BGP 数据库同步之后就使用更新消息交换路由信息。更新消息用来向邻居通告一条可行（Feasible）路由信息，或者撤销多条不可行的路由信息。一条消息中的所有属性都是关于特定的那条路径的。更新消息可以同时通告一条可行路由和撤销多条不可行路由。

更新报文结构如图 7-6 所示。

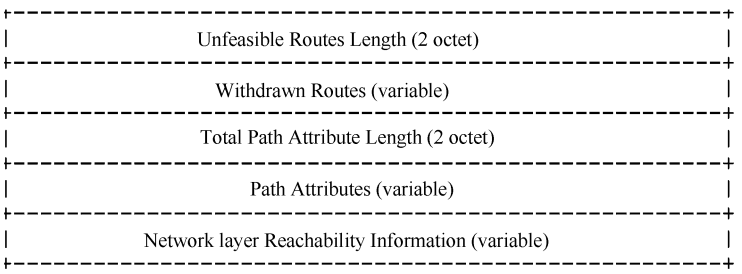

图 7-6　Update 报文结构

除了固定的报文头外，更新消息可以包含下面的字段：

- 不可行路由长度（2 字节）——指示撤销路由字段的长度。
- 撤销路由（变长）——如果有的话，其值为将被撤销的路由的 IP 地址前缀列表。
- 路径属性总长（2 字节）——指示路径属性字段的长度。
- 路径属性（变长）——相关路径的属性，包括属性类型、属性长度和属性值。有关属性的内容请参考 7.6 节。
- 网络层可达性信息（变长）——该字段包含了可达的 IP 地址前缀列表。

（3）存活消息

发送 Keepalive 消息是为了确认 BGP 对等体之间的连接是否依然存在，也是为了维护 BGP 邻居关系。对等体间定期交换 Keepalive 消息以使保持时间不超时，默认情况下，Keepalive 消息的周期为 60 秒。Keepalive 消息只有报文头。

（4）通知消息

当检测到一个错误条件时，路由器发送通知消息。通知消息发送后，路由器关闭 TCP 连接。通知消息包括 1 字节长的错误代码、1 字节长的错误子代码和变长的用来诊断错误的数据。

下面列出的是错误代码编号及其含义。

- 1：Message Header Error。
- 2：OPEN Message Error。
- 3：UPDATE Message Error。
- 4：Hold Timer Expired。

- 5：Finite State Machine Error。
- 6：Cease。

提示：BGP 消息数据包的长度为 19～4 096 字节。

7.5　BGP 邻接状态

BGP 使用“有限状态机”机制运行，其状态包括：

- 空闲（Idle）状态——该状态是 BGP 的最初状态。
  - 在空闲状态下，BGP 拒绝所有进入的 BGP 连接。在此状态下，由系统或操作员启动的“开始事件（Start Event）”发生后，系统初始化与其他 BGP 邻居的连接，同时也监听由其他邻居初始化的 TCP 连接。
  - 如果 BGP 探测到错误，它会关闭连接，并转入“空闲”状态。
- 连接（Connect）状态——该状态表明 BGP 正等待 TCP 连接的完成。
  - 如果连接成功，进入 OpenSent 状态；
  - 如果连接失败，进入 Active 状态。
- 主动（Active）状态——在该状态下，BGP 正在发起一个 TCP 连接，试图发现一个对等体。
  - 如果 TCP 连接成功，BGP 向对方发送 Open 消息，完成 TCP 初始化；否则再次向对方发起 TCP 连接，同时监听可能来自对等体的 TCP 会话，并把状态转变到 Connect。
  - 如果 BGP 检测到对等体正在试图与它建立连接，但是对等体的 IP 地址不是它期望的地址，那么它将拒绝连接，并继续监听可能来自对等体的 TCP 会话，仍然停留在 Active 状态。
- 打开发送（OpenSent）状态——在该状态下，BGP 等待来自对方的 Open 消息。
  - 当收到 Open 消息后，它检查所有字段的正确性，如果消息报文头或数据有错误，或者连接发生了碰撞，BGP 发出“通知”消息，并把状态转变到“空闲”。
  - 如果没有错误被检测出来，BGP 发送 Keepalive 消息，协商 Keepalive 和 Hold time 计时器，最终把状态转变到 OpenConfirm。
  - 如果收到来自底层协议的中断通知，BGP 将关闭连接并把状态转变到 Active。
- 打开确认（OpenConfirm）状态——该状态表示 BGP 正在等待 Keepalve 或 Notification 消息。
  - 如果收到 Keepalive 消息，进入 Established 状态。
  - 如果收到 Notification 消息，进入“空闲”状态。
- 已建立（Established）状态——邻居关系已经建立起来。
  - 在该状态下，BGP 可以交换“更新”消息、“通知”消息和 Keepalive 消息。
  - 如果在该状态下收到“通知”消息，或者“更新”消息有错误，或者计时器过期，BGP 则转入“空闲”状态。
  - 当连接状态从 Established 转入 Idle 后，所有源自该连接的路由条目都将被删除。

7.6　路 径 属 性

BGP 在路由信息中包含一系列描述路径信息的属性，这些属性用作选择最佳路径的尺度，称为路径属性。

7.6.1　属性分类

BGP 使用的属性（Attribute）可以分为如下几种：

- 公认（Well-known）属性——是所有运行 BGP 的路由器都必须识别的属性。
- 任选（Optional）属性——不必被所有运行 BGP 的路由器都支持的属性。
- 必遵（Mandatory）属性——必须出现在路由信息中的属性。
- 自决（Discretionary）属性——不必一定出现在路由信息中的属性。
- 可传递（Transitive）属性——可以传递给邻居的属性。
- 非传递（Non-transitive）属性——不能传递给邻居的属性。
- 不完全（Partial）属性——不被当前路由器支持的任选可传递属性，需要标记为该属性。

在使用时，这些属性是组合使用的，它们的有效组合如下：

- 公认的、必遵的属性（公认必遵属性）——公认必遵的属性必须出现在路由信息中。
- 公认的、自决的属性（公认自决属性）——公认自决的属性不是必须出现在路由信息中。
- 任选的、可传递的属性（任选可传递属性）——任选属性不要求所有的 BGP 路由器都支持，它可能是一个私有属性；如果它被支持的话，就可以把它传递给 BGP 邻居。不被当前路由器支持的任选可传递属性应该被继续传递给其他 BGP 邻居，但是该属性应该被标记为“不完全”属性。
- 任选的、非传递的属性（任选非传递属性）——任选非传递属性必须被不支持该属性的路由器删除。

7.6.2　BGP 属性

BGP 定义的属性包括以下几种。

① 公认必遵属性：

- AS 路径（AS Path）。
- 下一跳（Next Hop）。
- 起源（Origin）。

② 公认自决的属性：

- 本地优先（Local Preference）。
- 原子聚合（Atomic Aggregate）。

③ 任选可传递属性：

- 聚合者（Aggregator）。

- 团体（Community），Cisco 定义的属性。

④ 任选非传递属性：

- 多出口鉴别器（Multi-Exit-Discriminator，MED）。
- 发起者标识（Originator ID），Cisco 定义的属性。
- 集群列表（Cluster List），Cisco 定义的属性。

⑤ 权重（Weight）：Cisco 私有属性，仅在路由器本地有效，不传递给任何 BGP 邻居。

发起者标识属性和集群列表属性与路由反射器有关，这两个属性将在路由反射器一节中介绍，其他属性在本节介绍。

1. AS 路径属性

AS 路径属性描述到达目标网络所经由的所有 AS 的列表。当路由更新经过一个 AS 时，该 AS 号就被放在 AS 列表的前面，该列表的最后是路由信息起始 AS 的 AS 号。以图 7-7 所示的网络环境为例，解释 AS 路径列表是如何建立的。

AS 64500 中的路由器 A 向其邻居——AS 65500 中的路由器 C，通告网络 192.168.1.0，当这条路由信息到达路由器 C 时，路由器 C 把自己的 AS 号“前缀”（放在前面）放到这条路由信息的 AS 路径列表中。当 192.168.1.0 的路由信息到达路由器 B 时，它带有两个 AS 号。从路由器 B 的角度来看，到达网络 192.168.1.0 的路径是 65500→64500。

同理，路由器 A 到达网络 192.168.2.0 的 AS 路径是 65500→65000。

AS 路径属性在避免出现路由环路中起着十分重要的作用。如果 BGP 路由器收到一条路由信息，在该路由信息的 AS 路径属性值（AS 路径列表）中有自己的 AS 号，那么它将不接收该路由信息。

AS 号只在向 EBGP 邻居通告路由时被前缀到 AS 路径属性值中，在向 IBGP 邻居通告路由信息时不改变 AS 路径属性值。

2. 下一跳属性

下一跳属性用于指示到达目标网络的下一跳 IP 地址，也就是说，如果路由器收到一条路由信息，那么，通过读取路由信息中的下一跳属性值，该路由器就能知道，使用这条路由发送数据的下一跳 IP 地址。当前的 Internet 只采用一跳一跳的路由模式，知道下一跳 IP 地址对于路由器来说是至关重要的，因为对于路由器来说，知道了下一跳 IP 地址就等于有了到达目标网络的路由。

对于 EBGP，下一跳就是路由器的发送路由信息的接口 IP 地址，如图 7-8 所示。

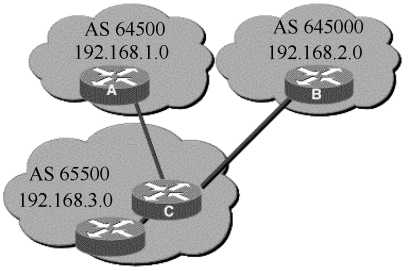

图 7-7　AS 路径属性

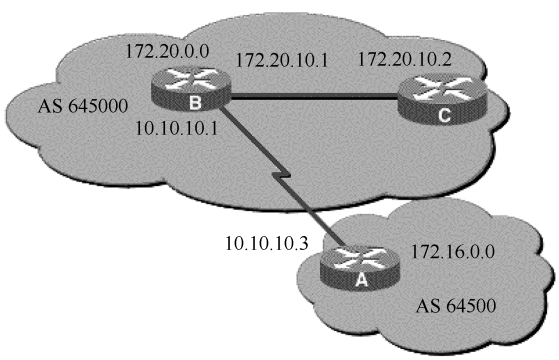

图 7-8　下一跳属性

路由器 A 将使用 10.10.10.3 作为下一跳属性值向路由器 B 通告网络 172.16.0.0，而路由器 B 将使用 10.10.10.1 作为下一跳属性值向路由器 A 通告网络 172.20.0.0。因此，路由器 B 使用下一跳地址 10.10.10.3 向网络 172.16.0.0 发送数据，而路由器 A 使用下一跳地址 10.10.10.1 向网络 172.20.0.0 发送数据。

对于 IBGP，路由器则把来自 EBGP 的下一跳属性值不做更改地传递给它的 IBGP 邻居。根据这一规则，路由器 B 将使用 10.10.10.3 作为下一跳属性值向路由器 C 通告网络 172.16.0.0。路由器 C 得到的信息是：要想把数据发送到网络 172.16.0.0，它必须把数据发送到 10.10.10.3。而图中显示的是，路由器 C 应该把数据发送给路由器 B，然后由路由器 B 再发送给路由器 A，最终到达目标网络。这意味着 10.10.10.3 的地址对于路由器 C 来说是不可达的。

因此，目前最重要的一点是，路由器 C 如何知道下一跳可达。这里有三种方法，一种方法是在路由器 B 和 C 之间运行某种 IGP 协议；另一种方法是在路由器 C 上配置一条静态路由，静态路由的下一跳地址为 172.20.10.1，即路由器 B；第三种方法是路由器 B 在向路由器 C 通告 172.16.0.0 的路由信息时，把下一跳属性的值改为自己的接口地址，即 172.20.10.1。总之，这些方法的目的是相同的，即让路由器 C 知道下一跳可达。否则，路由器 C 将不采纳这条路由信息，因为下一跳地址不可达。路由器 C 将没有到达 172.16.0.0 的路由。

下一跳属性在多路访问网络中的表现与上述情况会有些不同。

在广播型多路访问网络（如以太网）中运行 BGP 时，BGP 路由器将使用第三方 IP 地址作为下一跳属性的值，这样可以避免不必要的额外一跳，如图 7-9 所示。

在图 7-9 所示的网络中，路由器 A、B 和 C 与一个广播型多路访问网络相连。路由器 B 和 C 运行着某种 IGP，路由器 B 有到达网络 172.30.0.0 的路由。路由器 B 和路由器 A 运行着 BGP。当路由器 B 向路由器 A 通告关于 172.30.0.0 的路由信息时，它将使用 10.10.10.2 作为下一跳属性的值，而不是自己的 10.10.10.1 的 IP 地址。因为这 3 台路由器与一个广播型多路访问网络相连，对于路由器 A 来说，10.10.10.2 的下一跳地址是可达的。路由器 A 使用路由器 C 作为到达网络 172.30.0.0 的下一跳，比使用路由器 B 更直接，而且避免了无意义的额外一跳。

如果路由器之间的共享网络是非广播型多路访问（NBMA）介质，情况就不同了。假设路由器 A、B 和 C 通过帧中继连接，如图 7-10 所示，情况又会是怎样呢？

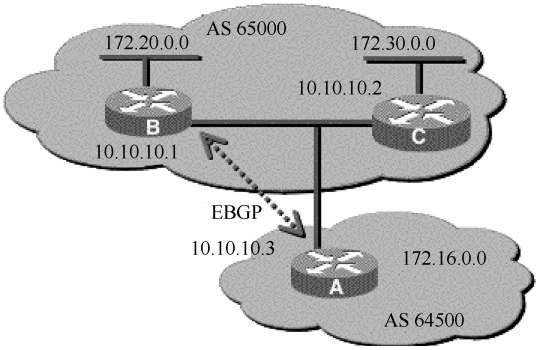

图 7-9　下一跳属性在 BMA 网络中

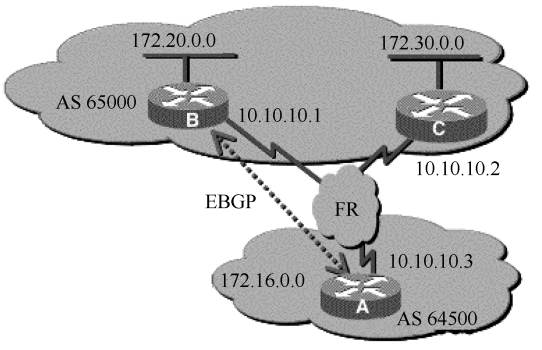

图 7-10　下一跳属性在 NBMA 网络中

图中路由器 B 和 A 之间存在虚链路，并且它们也建立了 EBGP 邻居关系。路由器 B 和 C 之间存在虚链路，并且它们之间运行着某种 IGP。路由器 A 和 C 之间不存在虚链路。

当路由器 B 通过 BGP 向路由器 A 通告 172.30.0.0 的路由信息时，使用 10.10.10.2 作为下一跳属性的值。当路由器 A 收到该路由信息时，由于路由器 A 没有到达 10.10.10.2 的路径，即

下一跳不可达，所以路由器 A 并不采纳该路由信息。路由器 A 的路由表里没有到达网络 172.30.0.0 的路由。

如果让路由器 A 采纳该路由，可以配置路由器 B 使用自己的地址（10.10.10.1）作为下一跳属性的值向路由器 A 通告 BGP 路由信息。

3. 本地优先属性

本地优先属性只在同一 AS 的路由器之间交换，当一个 AS 有多条与其他 AS 相连的链路时，它为 AS 内的路由器指示哪条链路是优先选择的出口路径。

本地优先值在路由器上配置，与配置较高本地优先值的路由器相连的路径将被优先使用。图 7-11 解释了本地优先属性的作用。

如图 7-11 所示，AS 64500 有两条出口链路，它可以从这两个方向上接收到网络 172.16.0.0 的路由信息。路由器 A 设置 172.16.0.0 的路由信息的本地优先属性值为 200。路由器 B 设置 172.16.0.0 的路由信息的本地优先属性值为 150。由于本地优先属性只在本 AS 内的 BGP 路由器之间传播（这里“本地”的含义是指本 AS），所以，AS 64500 内的 BGP 路由器都知道路由器 A 的本地优先属性值最高，如果向网络 172.16.0.0 发送数据，路由器 A 是优先选择的出口。

本地优先属性值在路由器上的默认值是 100。本地优先属性不传递给 EBGP 邻居。

4. MED 属性

MED 属性值也称为度量值，用于指示外部邻居（EBGP 邻居）进入本 AS 的优选路径。具有较低 MED 值的路径被优先选用。

MED 属性是一种让一个 AS 影响另一个 AS 如何选择路径的方法。因此，MED 属性与本地优先属性不同，MED 属性在自治系统之间交换，即在 EBGP 邻居之间交换，并且仅在相邻的 AS 间交换。MED 值不会被传递到第三个 AS 内，当路由信息被传递到第三个 AS 时，度量值将被设置为默认的 0 值。

默认情况下，路由器只比较来自同一个 AS 的不同 EBGP 邻居通告的同一个路由信息的度量值。如图 7-12 所示，路由器 B 将 MED 属性值设置为 150，路由器 C 将 MED 属性值设置为 200。当路由器 A 接收到来自同一个 AS（AS 65500）的不同邻居（路由器 B 和 C）通告的同一个路由信息（172.20.0.0）时，它选择 MED 值小的路由器（路由器 B）作为到达网络 172.20.0.0 的下一跳。

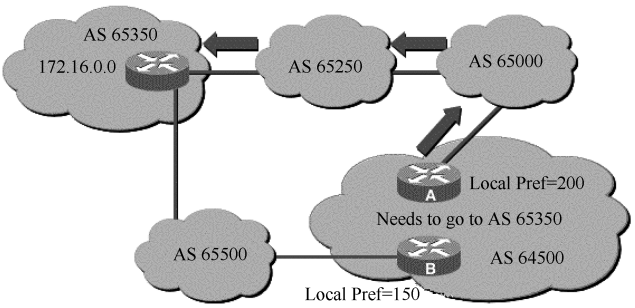

图 7-11　本地优先属性的作用

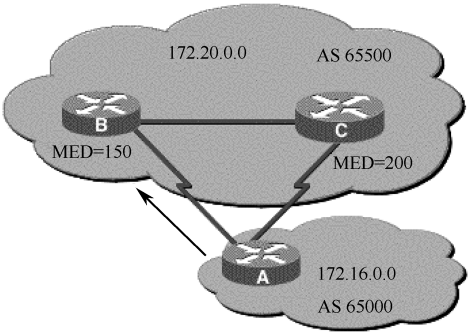

图 7-12　度量值的作用

如果让路由器比较来自不同自治系统的邻居的度量值，需要使用命令 bgp always-compare-med，但有可能出现路由环路。

5. 起源属性

起源属性指示路由信息的来源，起源属性值可以是如下 3 个值之一。

- i——表示路由信息来源于 IGP。使用“network”命令把 IGP 路由通告到 BGP 进程中的路由即是这种路由起源属性。（network 命令将在配置 BGP 部分讨论）
- e——表示路由信息起源与 EGP 协议，即是从 EGP 那里学习到的。这里的 EGP 协议不是前面提到的外部网关协议的总称，它是被 BGP 协议取代的一个早期外部网关协议。
- ?——表示路由的起源未知或通过某种其他方法学习到的，再被发布到 BGP 进程中的路由信息使用该属性值。

6. 团体属性

团体属性是把某些路由组织成一个团（体）并根据该属性应用路由判定的一种方法，类似于旅行社把一些散客组织成一个旅游团体。旅行社组团的最基本原则是这些客户的目的地是相同的，并且团队中的客户都具有某些共同的要求。类似地，团体也是一组共享某些共同属性的路由条目。不同于旅行社组团的是，每个路由条目可以属于多个团体，至于一条路由属于哪些团体是由网络管理员在路由器上根据需要设置的。路由器使用团体标志符对路由进行标记后，其他路由器可以根据这个团体标志符对路由信息进行控制，这样路由器可以对这一组路由进行控制而不是对组中每一条路由单独控制。

任一 BGP 路由器都可以在进入（Inbound）路由和外出（Outbound）路由中，或在进行再发布（Redistribution）路由时标记团体属性。任一路由器都可以根据团体属性值对进入的路由信息或外出的路由信息进行过滤，或者选择优先使用的路由。

每定义一个团体都分配它一个数字，这个数字就是该团体的团体标志符。团体标志符的范围是 1～4 294 967 200。除了用数字代表的自定义的团体外，一些公认的团体已经事先定义了：

- Internet——通告属于该团体的路由。
- No-export——属于该团体的路由不被通告到 EBGP 对等体。
- No-advertise——属于该团体的路由不被通告到任何对等体（IBGP/EBGP）。
- Local-as——即使在 BGP 邦联（Confederation）环境中，属于该团体的路由也不被通告到 EBGP 对等体。

默认地，所有路由都属于 Internet 团体。

团体属性可以在 AS 内和 AS 之间传递。BGP 路由器可以设置、附加和修改团体属性值。

如果让一台识别团体属性的路由器向它的对等体传递该属性，就必须对路由器进行人为配置，否则，团体属性将默认地从外出的 BGP 更新中被剥去。如果某台路由器不识别团体属性，则把该属性传递到下一台路由器。

提示：团体属性也称为 BGP 开销团体（BGP Cost Community）。由于很少使用 BGP 邦联，所以本书不讨论它。

7. 原子聚合属性

原子聚合属性与路由聚合（Aggregate）有关，这里的路由聚合与以前学习的路由总结的含义相同，实际上是对 BGP 表中的路由条目进行合并。当路由器向邻居通告的路由条目是聚合的路由条目时，原子聚合属性被设置，用来通知邻居这是一条聚合的路由条目，并表示原始信息可能已经丢失。

默认地，聚合路由被认为是始发于执行聚合的自治系统，并且聚合前的路由的 AS 号不被列出。

8. 聚合者属性

聚合者属性指出了进行路由聚合的 BGP 路由器 ID 及其所属 AS 号。

9. 权重属性

权重属性是 Cisco 私有属性，它针对每个不同的 BGP 邻居被本地化地配置在路由器上。权重属性只在本路由器上有效，不能被传递给任何 BGP 邻居。

权重属性值的范围是 0～65 535。起源于本路由器的路由（在本路由器上用 network 命令通告的或重再发布的路由，或者本地聚合的路由）的权重值为 32 768，其他路由的默认权重值为 0。

当一台路由器存在多条到达同一目的地的路由时，权重值高的路由被优先选用，如图 7-13 所示。

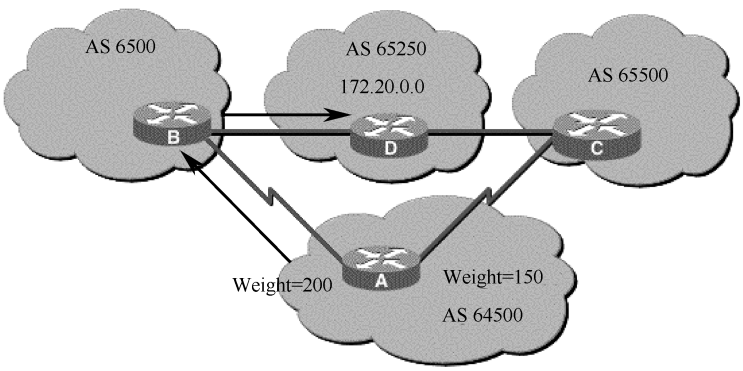

图 7-13　权重属性的作用

路由器 A 从路由器 B 和路由器 C 两个邻居处学到了 172.20.0.0 的路由信息，当需要发送数据到网络 172.20.0.0 时，它必须决定使用哪条路由。为此，路由器 A 为邻居 B 通告来的路由设置权重 200，为邻居 C 通告来的路由设置权重 150。这样，路由器 A 将使用路由器 B 作为到达网络 172.20.0.0 的优选下一跳。

小知识

BGP 属性类型编码:

- 类型编码 1——起源属性;
- 类型编码 2——AS 路径属性;

- 类型编码 3——下一跳属性;
- 类型编码 4——MED 属性;
- 类型编码 5——本地优先属性;
- 类型编码 6——原子聚合属性;
- 类型编码 7——聚合者属性;
- 类型编码 8——团体属性（Cisco 定义）;
- 类型编码 9——发起者 ID 属性（Cisco 定义）;
- 类型编码 10——集群属性（Cisco 定义）。

7.7　BGP 同步规则

BGP 同步（BGP Synchronization）规则要求：BGP 路由器不应该使用或向 EBGP 邻居通告从 IBGP 邻居那里学习到的 BGP 路由信息，除非该路由是本地的或者该路由存在于 IGP 数据库，即该路由也能从 IGP 学习到。同步规则要求 BGP 数据库与 IGP 数据库同步。这么做是为了使 AS 内的所有路由器都能达到同步，确保能够转发其他 AS 因使用该 AS 通告的路由而发送到该 AS 的那些数据。BGP 同步规则也保证了整个 AS 中路由信息的一致性，避免了 AS 内的路由“黑洞”。

下面以图 7-14 所示的网络环境为例说明同步开启时各 BGP 路由器的状态。

路由器 A 和 B 都运行 BGP 协议，并形成了 IBGP 邻居关系。在 AS 65500 中没有运行 IGP，路由器 A 和 B 上也没有任何静态路由条目。路由器 F 运行 BGP 协议，并和路由器 B 形成了 EBGP 邻居关系。路由器 E 运行 BGP 协议，并和路由器 A 形成了 EBGP 邻居关系。路由器 A 和路由器 B 共同构成了一条 AS 65000 和 AS 64500 之间的路由信息的转接路径。

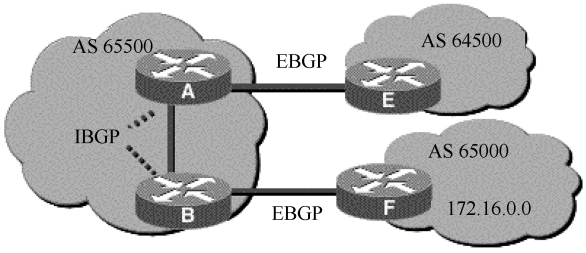

图 7-14　BGP 同步示例

路由器 F 在 BGP 进程中通告了网络 172.16.0.0，在同步开启时，情况如下：

- 路由器 F 在 BGP 进程中通告网络 172.16.0.0;
- 路由器 B 从 EBGP 邻居那里学习到 172.16.0.0 的路由信息;
- 路由器 B 把 172.16.0.0 的路由信息写入 BGP 库并且使用它，然后把该路由放置在路由表中;
- 路由器 B 向它的 IBGP 邻居路由器 A 通告 172.16.0.0 的路由信息;
- 路由器 A 学习到 172.16.0.0 的路由信息，写入 BGP 库但并不使用它，因为路由器 A 并没有从 IGP 那里学习到该路由信息，即不同步，路由表中也没有到达网络 172.16.0.0 的路由;

- 路由器 A 不把 172.16.0.0 的路由信息通告给它的 EBGP 邻居——路由器 E。

BGP 同步规则默认是开启的，可以人工关闭它。但是，只有当转接路径上的所有路由器都运行 BGP 时，关闭同步功能才是安全的。

依然使用图 7-11 所示的环境，当同步关闭时，情况如下：

- 路由器 F 在 BGP 进程中通告网络 172.16.0.0；
- 路由器 B 从 EBGP 邻居那里学习到 172.16.0.0 的路由信息；
- 路由器 B 把 172.16.0.0 的路由信息写入 BGP 库并且使用它，然后把该路由放置在路由表中；
- 路由器 B 向它的 IBGP 邻居路由器 A 通告 172.16.0.0 的路由信息；
- 路由器 A 学习到 172.16.0.0 的路由信息，写入 BGP 库但并不使用它，因为路由器 A 没有可达的下一跳（路由器 F 的接口地址），路由表中没有到达网络 172.16.0.0 的路由；
- 路由器 A 不把 172.16.0.0 的路由信息通告给它的 EBGP 邻居——路由器 E。

综上所述，如果让 BGP 路由器在不违反 BGP 同步规则的前提下，使用并向 EBGP 邻居通告从 IBGP 邻居学习到的路由信息，可以总结为如下两句话：

- 同步开启，BGP 和 IGP 必须同步，且有可达的下一跳（在运行了 IGP 的情况下，一般来说就能保证下一跳可达）；
- 同步关闭，必须有可达的下一跳。

提示：在 IOS 12.2(8)T 及以后的版本中，同步功能默认是关闭的；在此之前的版本中，同步功能默认是开启的。

7.8 路 由 选 择

路由器转发数据的依据是路由表。路由基本原理中有关路由器是如何选择路由（建立路由表）是这样描述的：路由器只把最可靠路由协议（管理距离值越小越可靠）的最佳路由（度量值越小越佳）写入路由表中。最佳路由也可能有多条。路由器为每一个协议堆栈建立一个路由表，比如，IP 路由表和 IPX 路由表是独立的两个表。

和 IGP 路由协议一样，BGP 路由协议也遵循这个基本原理。如果路由器上既运行了某种 IGP，同时又运行着 BGP，并且这两个协议都有到达同一个目的地的路由信息，那么，路由器首先通过比较管理距离来决定使用哪个路由协议的路由信息。如果 BGP 获胜，接下来路由器从多条到达同一目的地的 BGP 路由信息中选择一条（或几条）最佳路由写入路由表。

7.8.1 不予考虑的路由

对于一个特定的前缀（Prefix），在 BGP 路由器收到的所有路由条目中，有些路由不能作为最佳路径的候选者，在 BGP 库中，这些路由的前面没有有效标志（Valid Flag）。

路由器将忽略匹配如下条件的路由：

- 同步功能开启，但路由不同步。如果路由来自 OSPF，那么，OSPF 的路由器 ID 必须与 IBGP 邻居的路由器 ID 匹配。

- 下一跳不可达。
- 路由来自 EBGP 邻居，但 AS 路径列表中包含本地 AS 编号。
- 如果启用了 BGP Enforce-first-as 功能，但是邻居的 AS 编号在 AS 序列中并不是第一位。
- 策略（Policy）拒绝的路由。

7.8.2　选择最佳路由

在有效的最佳路径候选条目中，BGP 指定第一个有效路由为最佳路由，然后与其他候选路由比较，直到选出真正的最佳路由，把它写入路由表。

在 Cisco 路由器上，BGP 选择最佳路由的先后顺序如下：

⓪ 选择掩码最长的路由。

① 如果掩码长度相等，优先选用具有最高权重属性值的路由。

② 如果权重属性值相同，比较本地优先属性值。优先选用具有最高本地优先属性值的路由。如果路由不含本地优先属性，则认为它的本地优先属性值等于本地配置的值，或默认地认为等于 100。

③ 如果本地优先属性值相同，优先选用本路由器上使用 network 命令通告的路由，或本路由器再发布进 BGP 的路由，或者本地聚合的路由。前两者优先于后者。

④ 如果没有从本路由器始发的路由，比较 AS 路径长度。优先选用具有最短 AS 路径的路由。

⑤ 如果 AS 路径相同，比较路由起源属性值。优先选用具有最低起源代码的路由（i < e < ?）。

⑥ 如果起源相同，比较 MED 属性值。优先选用具有最低 MED 属性值的路由。

⑦ 如果 MED 属性值相同，优先选用 EBGP 路由，其次选用 IBGP 路由。如果存在多条 EBGP 最佳路由，转到⑨。

⑧ 如果都是 IBGP 路由，比较到达 BGP 下一跳地址的 IGP 路径 Metric，具有最低 Metric 值的路由被优先选用。如果存在多条最佳路由，继续下一步。

⑨ 确定是否需要安装多条 BGP 路由到路由表中。如果不需要，对于 EBGP 路由，继续下一步；对于 IBGP 路由转到⑪。

⑩ 如果都是 EBGP 路由，优先选用最先收到的路由（最老的那条）。这样做的目的是减小路由翻动（Route Flap）的影响。

⑪ 比较 BGP 邻居的路由器 ID 值。优先选用路由器 ID 值最低的路由。

⑫ 如果路由器 ID 值相同，使用集群列表长度最短的路由。

⑬ 比较邻居的 IP 地址。IP 地址最低的路由被选中。比较使用的 IP 地址是 neighbor 命令中使用的地址（请参考配置部分）。

提示：

① 这里总结的 BGP 路由器选择最佳路由的顺序不包括特殊情况，比如，人为使用了一些指令后，某个步骤可能会被跳过或者会插入一步。

② 运行 IOS 的路由器默认地为没有 MED 属性的路由指派 MED 值为 0，以使这种路由被优先选用。可以使用命令 bgp bestpath med missing-as-worst，使路由器为默认 MED 属性的路由指派一个最大值——4 294 967 294。

7.8.3　实现多路径支持

BGP 多路径（Multipath）特性允许在 IP 路由表中安装多条到达同一目的地的 BGP 路由条目，这些路径与最佳路径一同安装在路由表中进行负载分担。BGP 多路径特性不影响最佳路径的选择则，在数据库中，路由器仍然标记一条路由作为最佳路由，并把这条最佳路由通告给它的邻居。

BGP 多路径特性支持：

- EBGP 多路径——Maximum−paths *n*。
- IBGP 多路径——Maximum−paths ibgp *n*。
- eiBGP 多路径——Maximum−paths eibgp *n*。

BGP 最多安装 *n* 条最先收到的候选路由（Candidate Path）到路由表中。当前 *n* 的最大值为 6。默认值为 1，即不支持多路径负载分担。

提示：根据 IOS 版本的不同，以上命令及 *n* 的最大值会有所不同。

成为多路径候选条目的路由，其某些属性值需要与最佳路由的路径属性值相等，这些属性是：

- 权重。
- 本地优先。
- AS 路径长度。
- 起源。
- MED。
- 下列条件之一，
  - 邻接 AS 或 sub-AS（在 eiBGP 多路径附加要求之前）；
  - AS 路径 （在 eiBGP 多路径附加要求之后）。

一些多路径特性还为候选路由附加了要求。

为 eBGP 多路径候选路由附加的要求如下：

- 路由应该来自外部 BGP 邻居或邦联外部邻居（Confederation-External Neighbor）。
- 到达 BGP 下一跳的 IGP 度量应该与最佳路径到达 BGP 下一跳的度量相等。

为 iBGP 多路径候选路由附加的要求如下：

- 路由应该来自 IBGP 邻居。
- 到达 BGP 下一跳的 IGP 度量应该与最佳路径到达 BGP 下一跳的度量相等，除非路由器配置了不等值 IBGP 多路径负载分担。

提示：本书不讨论 BGP 邦联。 sub-AS 与 BGP 邦联有关，读者可以自行查阅 Cisco 文档。

7.9　BGP 路由反射

同步规则是控制 BGP 运行行为的重要规则，控制 BGP 行为的另一规则是 BGP 的 Split Horizon 规则。

提示：有些书籍中称其为“横向隔离规则”、“水平隔离规则”、“水平分割规则”。作者认为这种翻译是不正确的。Horizon 是“空间”、“地平线”的意思。在 BGP 中，Split Horizon 用来描述对路由信息传播范围的限制。由于作者水平有限，也没有找到一个合适的中文术语对应它，因此，本书中直接引用英文术语。

7.9.1　Split Horizon

BGP 的 Split Horizon 规则规定：路由器通过 IBGP 对等体学习到的路由不能被传输到其他 IBGP 对等体。如图 7-12 所示，路由器 A 通过 IBGP 从路由器 B 那里学习的路由，永远不会被传递给路由器 C。这样，路由器 A 的路由信息的扩散空间就被“Split”了，使其不能扩展得更大，比如扩展到路由器 C。

BGP 的这个规则对于确保路由不会在 AS 内产生环路是很重要的，这与距离矢量型路由协议使用 Split Horizon 规则的目的相同。但是，这样就需要同一 AS 内的 IBGP 邻居之间建立全互连的 IBGP 邻接关系。以图 7-15 所示拓扑为例，路由器 B 和 C 之间也需要建立 IBGP 邻接关系，这样就是全互连的 IBGP 邻接关系，路由器 C 可以直接从路由器 B 那里获得路由信息。

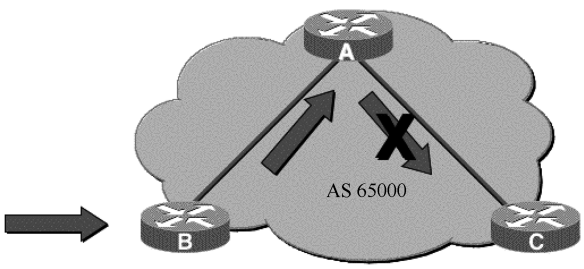

图 7-15　Split Horizon

全互连的 IBGP 邻接关系是不易扩展的。随着路由器数量的增加，所需的 TCP 会话数量也大大增加。当只有 13 台路由器时，会话数就有 78（= 13×(13−1)/2）个之多，如图 7-16 所示。

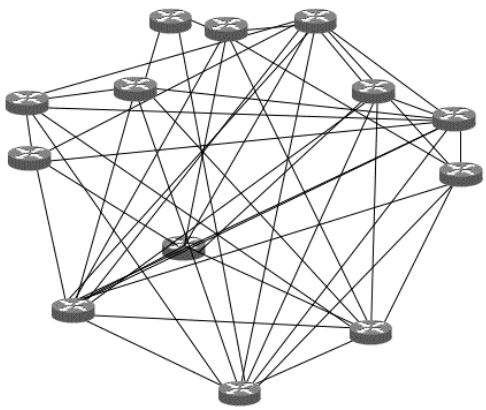

图 7-16　全互连的 IBGP

除了 TCP 会话数量大外，路由信息数据的总量也是问题。当路由信息在各个 BGP 对等体

之间被传递时，可能会在某些链路上被复制多次（与具体拓扑有关）。如果一个大型 AS 的物理拓扑包括一些 WAN 链路，那么 IBGP 会话可能会占用这些链路的大量带宽。

解决这些问题的方法是使用路由反射器。

7.9.2 路由反射器

路由反射器（Route Reflector）技术让配置为路由反射器的路由器向其他 IBGP 对等体传递（反射）从 IBGP 邻居学习到的路由信息。该技术弥补了由 BGP Split Horizon 规则带来的局限性，减少了 BGP 的 TCP 会话数量，也减少了 BGP 路由信息传播量。

配置了 BGP 路由反射器，就不再需要全互连的 IBGP 对等体。路由反射器不影响 IP 数据包所要经过的路径，只与通告路由信息有关。但是，规划和配置路由反射器也需要小心，如果路由反射器配置不正确，可能会产生路由环路，这就会影响到 IP 数据包的发送。

在 AS 内可以有多个路由反射器，既是为了冗余也是为了分组，以达到进一步减少 IBGP 会话数量和路由信息的复制量。

配置路由反射器只需要很少的配置命令和步骤，因此不必一次完成所有的配置工作，而且，配置为路由反射器的路由器和不是路由反射器的路由器可以共存。

7.9.3 路由反射器术语

- 路由反射器——能够把它通过 IBGP 学习到的路由信息通告（反射）给其他 IBGP 邻居的路由器。
- 路由反射器客户（Client）——路由反射器反射路由的目标对象称为路由反射器的客户。路由反射器客户只需要与路由反射器建立 IBGP 会话，客户之间不需要 IBGP 对等关系。不是路由反射器客户的路由器称为非客户。
- 集群（Cluster）——路由反射器及其客户的集合称为集群。通常，一个集群只有一个路由反射器，路由反射器的路由器 ID 即为集群的标志号。为了避免单点失败故障，增加冗余性，集群也可能会有一个以上的路由反射器，在这种情况下，集群中的所有路由反射器都需要配置“集群 ID”。集群 ID 使路由反射器能够识别来自同一集群内的其他路由反射器的路由信息，有助于避免路由环路。

在众多 BGP 属性中，有两个属性与路由反射器有关，这两个属性是发起者标识（Originator ID）和集群列表（Cluster list）。

发起者标识属性由路由反射器创建，该属性携带路由始发者的路由器 ID。如果因为配置不当而导致路由信息又回到始发者，那么始发者将忽略它。

集群列表是路由信息所经过的集群的 ID 序列表。如果接收到的路由更新中包含自己的集群列表，那么路由器将忽略该更新。通过该属性，可以避免因配置不当造成的路由环路。

7.9.4 路由反射器的规划

当在 AS 内规划路由反射器时，根据具体的网络拓扑可以将 AS 划分成多个集群，每个集群中至少有一个路由反射器和数个客户。也可以考虑在一个集群中使用多个路由反射器。

路由反射器之间、路由反射器与非客户之间以及非客户之间必须建立全互连的 IBGP 对等关系，以确保路由信息能传遍整个 AS。图 7-17 是路由反射器规划的一个示例。

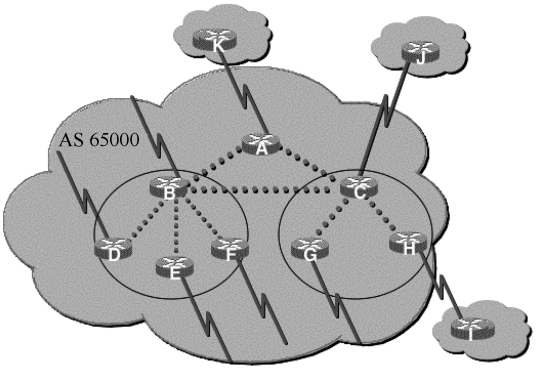

图 7-17　一个路由反射器规划示例

图中的虚线代表 IBGP 会话，折线代表 EBGP 会话。图中没有显示物理连接。

图中规划了两个集群，路由器 B、D、E 和 F 形成一个集群，路由器 B 是该集群的路由反射器。路由器 C、G 和 H 形成一个集群，路由器 C 是该集群路由反射器。路由器 A、B 和 C 建立了全互连的 IBGP 关系。

7.9.5　路由反射规则

当路由反射器接收到路由更新时，它反射路由的规则如下：

- 如果路由信息来自它的一个客户对等体，那么它把路由信息反射给所有非客户对等体以及它的所有客户对等体（除了信息源客户）。
- 如果路由信息来自一个非客户对等体，那么它把路由信息反射给集群中的所有客户对等体。
- 如果路由信息来自一个 EBGP 对等体，那么它把路由信息反射给集群中的所有客户对等体以及所有对等体。

当路由反射器客户接收到路由更新时，做如下处理：

- 如果路由信息来自它的路由反射器，那么它把路由信息通告给它的 EBGP 对等体（如果有的话）。
- 如果路由信息来自它的 EBGP 对等体，那么它把路由信息通告给它的路由反射器。

以图 7-17 所示的路由反射器规划为例，一种路由信息的反射情况可能是这样的：

- 路由器 H 收到路由器 I（EBGP 邻居）的路由更新信息，它把路由信息发送给路由器 C。
- 路由器 C 接收到路由器 H（客户）的路由更新信息，它把路由信息发送给路由器 G、A、B 和 J。
- 路由器 C 接收到路由器 A 或 B（非客户）的路由更新信息，它把路由信息发送给路由器 G 和 H 以及路由器 J。
- 路由器 C 接收到路由器 J（EBGP 邻居）的路由更新信息，它把路由信息发送给路由器 G 和 H 以及 A 和 B。

7.9.6　路由反射器设计技巧

当决定要使用路由反射器时，首先要考虑的是哪台路由器应该作为路由反射器，哪些路由器作为它的客户。在设计中应该考虑以下三点。

- 遵从网络的物理拓扑结构，这样可以保证数据包的转发不受影响。遵从物理拓扑结构，实际上是遵从数据的物理路径。应该把与路由反射器有直连链路的路由器作为它的客户，否则可能会导致路由环路。
- 一次配置一个路由反射器，然后把那些多余的 IBGP 会话拆除。
- 一个集群中配置一个路由反射器。

图 7-18 所示的设计是一个不遵从物理拓扑结构的案例。路由器 E 同时是路由反射器 C 和 D 的路由反射器客户。路由器 E 与路由器 C 相连的接口地址是 *x*，和路由器 D 相连的接口地址是 *y*。路由器 A 是路由反射器 D 的客户，路由器 B 是路由反射器 C 的客户。路由器 E 通告了网络 10.0.0.0/8 的路由信息。

这样的设计会导致下面的情况发生：

- 路由器 A 从路由器 D 那里学习到 10.0.0.0/8 路由信息，下一跳地址是 *y*。
- 路由器 B 从路由器 C 那里学习到 10.0.0.0/8 路由信息，下一跳地址是 *x*。
- 假设路由器 A 到达 *y* 的最佳路径是经由 B，那么路由器 A 将把目标地址为 10.0.0.0/8 的数据发送给路由器 B。
- 假设路由器 B 到达 *x* 的最佳路径是经由 A，那么路由器 B 将把目标地址为 10.0.0.0/8 的数据发送给路由器 A。
- 路由器 A 和 B 之间形成了环路。

如果遵从物理拓扑结构设计，上述路由反射器设计的案例应该如图 7-19 所示。

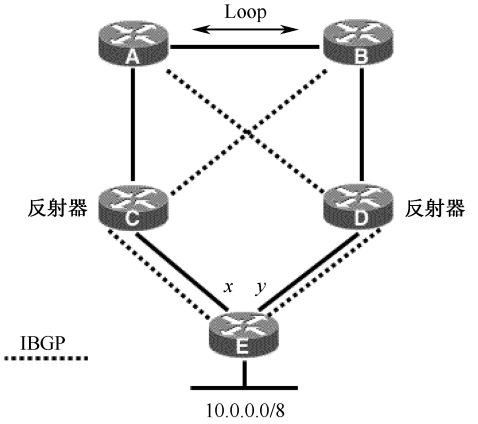

图 7-18　不遵从物理拓扑的路由反射器设计

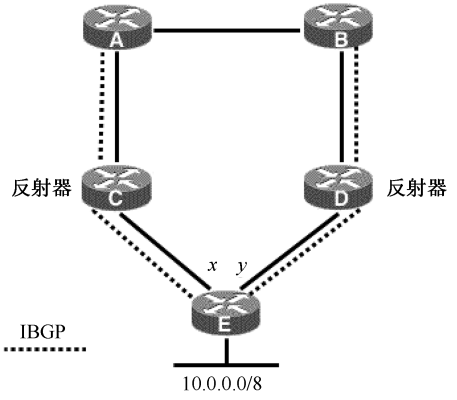

图 7-19　遵从物理拓扑的路由反射器设计

在这个遵从物理拓扑的路由反射器设计案例中，数据是这样流动的：

- 路由器 A 从路由器 C 那里学习到 10.0.0.0/8 路由信息，下一跳地址是 *x*。
- 路由器 B 从路由器 D 那里学习到 10.0.0.0/8 路由信息，下一跳地址是 *y*。
- 路由器 A 到达 *x* 的最佳路径是经由 C，那么路由器 A 将把目标地址为 10.0.0.0/8 的数据发送给路由器 C。路由器 C 把数据转发给路由器 E。

- 路由器 B 到达 *y* 的最佳路径是经由 D，那么路由器 B 将把目标地址为 10.0.0.0/8 的数据发送给路由器 D。路由器 D 把数据转发给路由器 E。

在这个遵从物理拓扑的路由反射器设计案例中没有形成环路。

7.10　配置 BGP 路由协议

7.10.1　基本配置命令

在路由器上启动 BGP 路由选择协议的基本命令需要两个，一个是起用 BGP 协议，另一个是配置对等体。

① 在路由器上起用 BGP 协议，命令语法如下：

```
router(config)# router bgp AS-number
```

- *AS-number*：本路由器所在的自治系统号。

② 配置对等体，命令语法如下：

```
router(config-router)# neighbor ip-address remote-as AS-number
```

- *ip-address*：邻居的 IP 地址。
- *AS-number*：邻居所在的自治系统号。

路由器根据自己所在的 AS 编号和邻居所在的 AS 编号，决定建立何种 BGP 邻接关系。如果在“neighbor…remote-as…”命令中的 AS 号与自己所在的 AS 号不同，那么路由器建立一个 EBGP 邻接关系；如果相同，则建立一个 IBGP 邻接关系。

EBGP 对等体之间一般需要有直连的物理链路，因此，BGP 默认地设置建立 EBGP 会话的数据包的 TTL 值为 1。对于 IBGP 对等体，它们之间则不必一定有直连的物理链路，因此，BGP 默认地设置建立 IBGP 会话的数据包的 TTL 值为 255。

要关闭一个已经存在的 BGP 会话，可以使用如下命令：

```
router(config-router)# neighbor ip-address shutdown
```

打开一个被关闭的会话，使用如下命令：

```
router(config-router)# no neighbor ip-address shutdown
```

需要删除一个邻居关系，使用如下命令：

```
router(config-router)# no neighbor ip-address remote-as AS-number
```

提示：EBGP 邻接关系也可以在非直连的路由器之间建立，但需要设置 TTL 值：router(config-router)# neighbor *ip-address* ebgp-multihop [*ttl*]。如果不声明 TTL 值，则默认使用 255。

7.10.2　环回接口与 BGP 会话

环回接口是一个软件意义上的虚拟接口，与硬件接口独立，因此它不受某个物理接口故障的影响。使用环回接口地址建立 BGP 会话，避开了因用于建立 BGP 会话的物理接口故障而造

成 BGP 会话中断的可能性。

如果路由器在“neighbor *ip-address* remote-as *AS-number*”命令中使用的 IP 地址是它邻居的某个环回接口的地址，那么必须在邻居路由器上增加额外的配置指令。让邻居路由器通知它的对等体，它使用的是环回接口发起的 TCP 连接，而不是物理接口。

命令及语法如下：

```
router(config-router)# neighbor ip-address update-source loopback loopback-interface-number
```

- *loopback-interface-number*：环回接口编号。

例如，在图 7-20 所示的网络中，路由器 A 使用物理接口与路由器 B 建立 BGP 会话；路由器 B 使用环回接口 0 与路由器 A 建立 BGP 会话，它们的配置如下所述。

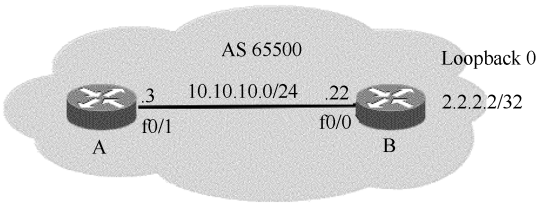

图 7-20　使用环回接口配置

① 配置路由器 A：

```
routerA(config)# router bgp 65500
routerA(config-router)# neighbor 2.2.2.2 remote-as 65500
```

② 配置路由器 B：

```
routerB(config)# router bgp 65500
routerB(config-router)# neighbor 10.10.10.3 remote-as 65500
routerB(config-router)# neighbor 10.10.10.3 update-source loopback 0
```

7.10.3　通告路由

在路由器上向 BGP 路由进程中通告路由信息的方法有以下几种：

- 把本地直连网络（直连路由）通告到 BGP；
- 把路由表中的 IGP 路由通告到 BGP；
- 把 IGP 再发布到 BGP；
- 通告聚合的 BGP 路由。

在路由器上通告路由意味着被通告到 BGP 的路由都起源于本路由器。一台运行 BGP 的路由器可以不通告任何路由，那么它只需要使用和转发其他路由器通告的 BGP 路由即可。

本节介绍前两种方法，这两种方法都使用“network”命令。后两种方法将在后续的章节中介绍。

用 network 命令通告路由的语法如下：

```
router(config-router)# network network-number mask
```

- *network-number*：被通告的网络号。
- *mask*：被通告的网络的子网掩码。

使用该命令通告路由的前提条件是：被通告的路由必须是 IP 路由表中的 IGP 路由（Mask

的长度也必须与路由表中的长度相同），或者是本路由器的直连路由。

该命令可以通告子网路由、主网路由、超网路由或者一个总结的 IGP 路由。

7.10.4　BGP 配置示例

1．配置 BGP

BGP 配置使用的拓扑结构如图 7-21 所示，网络中规划了 3 个自治系统。

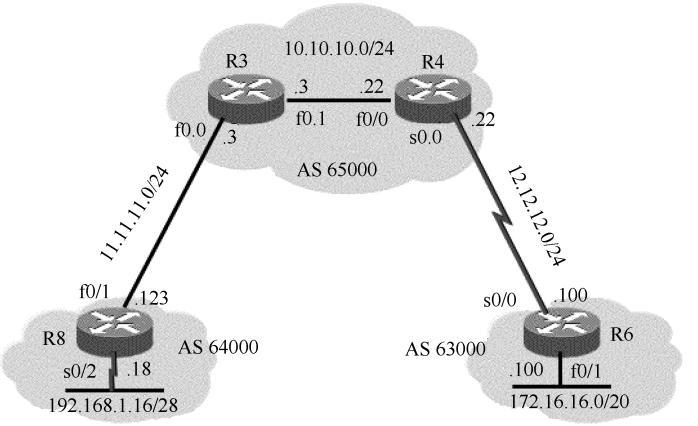

图 7-21　BGP 配置示例

路由器 R3 和 R4 在自治系统 65000 中，并且建立了 IBGP 邻接关系；路由器 R8 在自治系统 64000 中，并且与路由器 R3 建立了 EBGP 邻接关系；路由器 R6 在自治系统 63000 中，并且和路由器 R4 建立了 EBGP 邻接关系。

在 AS 65000 中运行的 IGP 是 RIP，在路由器 R6 上通告了网络 172.16.16.0/20，在路由器 R8 上通告了网络 192.168.1.16/28。

下面是这 4 台路由器的配置列表（只列出与 BGP 相关的配置）。

（1）路由器 R3 的配置

```
hostname "R3"
!
interface FastEthernet0/0
 ip address 11.11.11.3 255.255.255.0
 duplex auto
 speed auto
!
interface FastEthernet0/1
 ip address 10.10.10.3 255.255.255.0
 duplex auto
 speed auto
!
```

```
router rip
 network 10.0.0.0
 network 11.0.0.0
!
router bgp 65000
 bgp log-neighbor-changes
 neighbor 10.10.10.22 remote-as 65000
 neighbor 11.11.11.123 remote-as 64000
!
ip classless
ip route 172.16.16.0 255.255.240.0 10.10.10.22
```

（2）路由器 R8 的配置

```
hostname "r8"
!
interface FastEthernet0/1
 ip address 11.11.11.123 255.255.255.0
 duplex auto
 speed auto
!
interface Serial0/2
 ip address 192.168.1.18 255.255.255.240
 clockrate 64000
!
router bgp 64000
 bgp log-neighbor-changes
 network 192.168.1.16 mask 255.255.255.240
 neighbor 11.11.11.3 remote-as 65000
!
ip classless
```

（3）路由器 R4 的配置

```
hostname "r4"
!
interface FastEthernet0/0
 ip address 10.10.10.22 255.255.255.0
 duplex auto
 speed auto
!
interface Serial0/0
```

```
 ip address 12.12.12.22 255.255.255.0
 clockrate 64000
!
router rip
 network 10.0.0.0
 network 12.0.0.0
!
router bgp 65000
 bgp log-neighbor-changes
 neighbor 10.10.10.3 remote-as 65000
 neighbor 12.12.12.100 remote-as 63000
!
ip classless
ip route 192.168.1.16 255.255.255.240 10.10.10.3
```

（4）路由器 R6 的配置

```
hostname "R6"
!
interface Serial0/0
 ip address 12.12.12.100 255.255.255.0
!
interface FastEthernet0/1
 ip address 172.16.16.100 255.255.240.0
 duplex auto
 speed auto
!
router bgp 63000
 bgp log-neighbor-changes
 network 172.16.16.0 mask 255.255.240.0
 neighbor 12.12.12.22 remote-as 65000
!
ip classless
```

在 Cisco 路由器上，BGP 同步功能默认是开启的。本例使用的就是默认模式。本例不是用来讨论同步的，但已经考虑了同步的影响。有关同步的详细分析将在下一节讨论。

提示: 在路由器 R3 和 R4 上分别设置了一条静态路由，请读者考虑静态路由的作用是什么？

2．检验 BGP 运行

下面使用的命令主要用来检验 BGP 的运行，以前学习过的命令这里不再重复说明（但仍然在 BGP 环境下可用），除非为了解释 BGP 运行原理。

（1）查看 BGP 邻居

下面的输出来自路由器 R3（部分内容）：

```
R3#show ip bgp neighbors
1. BGP neighbor is 11.11.11.123,  remote AS 64000, external link
2. BGP version 4, remote router ID 192.168.1.18
3. BGP state = Established, up for 00:52:15
4. Last read 00:00:15, hold time is 180, keepalive interval is 60 seconds
……

BGP neighbor is 10.10.10.22,  remote AS 65000, internal link
  BGP version 4, remote router ID 12.12.12.22
  BGP state = Established, up for 00:42:15
  Last read 00:00:15, hold time is 180, keepalive interval is 60 seconds
…….
```

输出显示，路由器 R3 有两个 BGP 邻居，一个是 EBGP，另一个是 IBGP。

第 1 行显示用来建立 TCP 连接的邻居的 IP 地址、邻居所在的自治系统编号和链路类型。本链路是外部链路（external link），表示是 EBGP 邻接关系。

第 2 行显示 BGP 版本号，当前是 4，以及邻居路由器的 ID（remote router ID）。

第 3 行显示 BGP 邻接状态。输出显示当前状态是“已建立”，以及会话已经建立的时间。

第 4 行显示的是最近一次收到 keepalive 或 update 消息距目前的时间，以及保持时间和 keepalive 间隔时间。

第二个邻居的输出含义类似，这里不再赘述。其他路由器上的邻居建立情况这里也不再列出。

（2）查看 BGP 数据库

① 路由器 R4 的 BGP 库。

```
r4#show ip bgp
1. BGP table version is 7, local router ID is 12.12.12.22
2. Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
3. Origin codes: i - IGP, e - EGP, ? - incomplete

4.     Network          Next Hop            Metric  LocPrf  Weight  Path
*>   172.16.16.0/20     12.12.12.100         0               0      63000   i
*>i  192.168.1.16/28    11.11.11.123         0       100     0      64000   i
r4#
```

第 1 行显示了 BGP 表（数据库）的更新版本号以及本地路由器 ID。

第 2 行是描述路由条目状态（Status Codes）的代码。路由状态代码包括：

- s——Suppressed（被抑制路由），被聚合的路由标记为 s。
- d——Damped（消沉路由），Dampening 功能启动后，翻动（Flapping）频繁的路由条目被标记为消沉路由。

- h——History（历史路由），一旦路由翻动，则被标记 h。意思是根据历史信息路由器没有最佳路由。
- *——valid（有效路由）。
- >——best（最佳路由）。
- i——internal（来自 IBGP 对等体的路由）。

第 3 行是描述路由起源（Origin Codes）的代码。路由起源代码包括（请参考 BGP 属性一节）：

- i——IGP（起源于 IGP 的路由）。
- e——EGP（起源于 EBG 的路由）。
- ?——Incomplete（起源于未知的路由）。

第 4 行是数据库标题头，标题头下面列出的是具体路由信息。

具体路由信息部分包含 8 列，除最前一列和最后一列没有标题外，其他 6 列均有标题，这 8 列的含义是：

- 第 1 列代表路由状态。
- 最后一列代表路由起源属性值。
- 第 2 列代表路由前缀（Network）。
- 第 3 列代表下一跳（Next hop）属性值。
- 第 4 列代表度量（Metric）值，即 MED 值。
- 第 5 列代表本地优先（LocPrf）属性值。
- 第 6 列代表权重（Weight）属性值。
- 第 7 列代表 AS 路径（Path）属性值。

路由器 R4 的 BGP 数据库显示，它学习到了两个路由条目，这两个路由条目都是有效路由，并且是最佳路由。路由器只把最佳路由安装到路由表中，除非启用了不等值路径负载分担特性。因此，这两个路由条目可能会出现在路由表中（前提条件是 BGP 协议在与其他正在运行的协议或者静态路由比较管理距离时胜出）。数据库还显示，这两个路由条目都起源于 IGP，因为这两个路由条目是路由器 R6 和 R8 分别用 network 命令通告的。

另外，这两条路由的不同之处在于，一个是外部路由条目（来自 EBGP 邻居 R6），而另一个是内部路由条目（来自 IBGP 邻居 R3）。

再者，从数据库中还可以看出，路由器 R4 发送数据到达网络 172.16.16.0/20 只需经由 AS 63000 即可到达，使用的下一跳地址是 12.12.12.100（R6 的接口地址），MED 值为 0。因为该路由是外部路由（来自 EBGP 邻居），因此，本地优先和权重属性对它来说是没有意义的。

路由器 R4 发送数据到达网络 192.168.1.16/28 只需经由 AS 64000 即可到达，使用的下一跳地址是 11.11.11.123（R8 的接口地址），MED 值权重属性值都为 0。本地优先属性值是默认值 100。

② 路由器 R3 的 BGP 库。

```
R3# show ip bgp
BGP table version is 3, local router ID is 11.11.11.3
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete
```

```
     Network          Next Hop            Metric  LocPrf  Weight  Path
*>i  172.16.16.0/20   12.12.12.100             0     100       0  63000      i
*>   192.168.1.16/28  11.11.11.123             0               0  64000      i
R3#
```

③ 路由器 R6 的 BGP 库。

```
R6#sh ip bgp
BGP table version is 5, local router ID is 172.16.16.100
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
*> 172.16.16.0/20   0.0.0.0                  0               32768      i
*> 192.168.1.16/28  12.12.12.22                              0  65000 64000  i
R6#
```

从 R6 的数据库可以看出，本地始发的路由（172.16.16.0/20）的下一跳为 0.0.0.0，意思是本路由器可以直达；本地始发的路由的默认权重值为 32768。

从数据库中还可以看出，AS 路径的排列方法是路由经过的 AS 都把自己的 AS 号放在队列的最前边（最左边），即把自己的 AS 号前缀（Prepended）到 AS 路径中。

④ 路由器 R8 的 BGP 库。

```
r8#sh ip bgp
BGP table version is 13, local router ID is 192.168.1.18
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
*> 172.16.16.0/20   11.11.11.3                               0  65000 63000  i
*> 192.168.1.16/28  0.0.0.0                  0               32768      i
r8#
```

（3）查看路由表

① 路由器 R4 的路由表。

```
r4#sh ip route
Codes: C - connected, S - static, I - IGRP, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
       i - IS-IS, L1 - IS-IS level-1, L2 - IS-IS level-2, ia - IS-IS inter area
       * - candidate default, U - per-user static route, o - ODR
       P - periodic downloaded static route

Gateway of last resort is not set
```

```
        172.16.0.0/20 is subnetted, 1 subnets
B       172.16.16.0 [20/0] via 12.12.12.100, 00:09:42
        10.0.0.0/24 is subnetted, 1 subnets
C       10.10.10.0 is directly connected, FastEthernet0/0
R    11.0.0.0/8 [120/1] via 10.10.10.3, 00:00:17, FastEthernet0/0
        12.0.0.0/24 is subnetted, 1 subnets
C       12.12.12.0 is directly connected, Serial0/0
        192.168.1.0/28 is subnetted, 1 subnets
S       192.168.1.16 [1/0] via 10.10.10.3
r4#
```

路由表显示，R4 通过 BGP 学习到了 172.16.16.0/20 的路由条目；路由表还显示出 EBGP 路由的管理距离为 20。

② 路由器 R3 的路由表。

```
R3#sh ip route
(代码省略)

        172.16.0.0/20 is subnetted, 1 subnets
S       172.16.16.0 [1/0] via 10.10.10.22
        10.0.0.0/24 is subnetted, 1 subnets
C       10.10.10.0 is directly connected, FastEthernet0/1
        11.0.0.0/24 is subnetted, 1 subnets
C       11.11.11.0 is directly connected, FastEthernet0/0
R    12.0.0.0/8 [120/1] via 10.10.10.22, 00:00:16, FastEthernet0/1
        192.168.1.0/28 is subnetted, 1 subnets
B       192.168.1.16 [20/0] via 11.11.11.123, 00:05:32
R3#
```

③ 路由器 R6 的路由表。

```
R6#sh ip route
(代码省略)

        172.16.0.0/20 is subnetted, 1 subnets
C       172.16.16.0 is directly connected, FastEthernet0/1
        12.0.0.0/24 is subnetted, 1 subnets
C       12.12.12.0 is directly connected, Serial0/0
        192.168.1.0/28 is subnetted, 1 subnets
B       192.168.1.16 [20/0] via 12.12.12.22, 00:06:11
R6#
```

④ 路由器 R8 的路由表。

```
r8# sh ip route
(代码省略)
```

```
        172.16.0.0/20 is subnetted, 1 subnets
B       172.16.16.0 [20/0] via 11.11.11.3, 00:12:40
        11.0.0.0/24 is subnetted, 1 subnets
C       11.11.11.0 is directly connected, FastEthernet0/1
        192.168.1.0/28 is subnetted, 1 subnets
C       192.168.1.16 is directly connected, Serial0/2
r8#
```

提示：因为路由器 R3 和 R4 的同步功能是开启的，所以给它们设置了静态路由，目的是让路由器满足同步条件。设置静态路由只是在实验环境下使用的一种简单方法，并非唯一方法。在实际网络环境中，静态路由方法可能不合适，在这种情况下，需要考虑使用其他方法或关闭同步功能。

7.10.5　更改下一跳属性值

下一跳是公认必遵属性。对于 EBGP，下一跳就是路由器的发送路由信息的接口 IP 地址。对于 IBGP，路由器则把来自 EBGP 的下一跳属性值不做更改地传递给它的 IBGP 邻居。

从 7.10.4 节的 BGP 配置示例中可以看到，路由器 R3 通过 BGP 把 192.168.1.16/28 的路由信息通告给 R4，该路由的下一跳属性值是 11.11.11.123（R8 的接口地址），并不是路由器 R3 的接口地址 10.10.10.3。这样就带来一个问题，如果路由器 R4 使用这条路由的话，它必须知道如何到达 11.11.11.123，即下一跳可达。为了解决这个问题，在路由器 R4 上运行了 RIP 协议，保证了下一跳可达。

当然，同样的问题也会出现在路由器 R3 上。为了解决这个问题，在 R3 上也运行了 RIP 协议，保证了下一跳可达。

不运行 IGP 而又能保证下一跳可达的方法是更改下一跳属性值。

如果让路由器 R3 把通告给路由器 R4 的 BGP 路由的下一跳属性值更改为自己接口的地址，那么，不必运行 RIP 协议就能保证下一跳就可达，因为路由器 R3 和路由器 R4 是直连的。

把下一跳属性值更改为自己接口地址的命令语法如下：

```
router(config-router)# neighbor ip-address next-hop-self
```

- *ip-address*：邻居的地址。

① 在 R3 上的配置更改为（与本命令相关部分）。

```
router bgp 65000
 neighbor 10.10.10.22 remote-as 65000
 neighbor 11.11.11.123 remote-as 64000
 neighbor 10.10.10.22 next-hop-self
```

② 在 R4 上的配置更改为（与本命令相关部分）。

```
router bgp 65000
 neighbor 10.10.10.3 remote-as 65000
 neighbor 12.12.12.100 remote-as 63000
 neighbor 10.10.10.3 next-hop-self
```

7.10.6　BGP 同步分析

1. 同步开启但不满足同步条件

分析同步功能开启时各路由器的表现情况使用的拓扑如图 7-22 所示。路由器 R4、R5、R6 位于 AS 63000 中，并且构成 AS 64000 和 AS 65000 之间的转接路径。路由器 R4 和 R6 运行 BGP 并建立 BGP 邻接关系，路由器 R5 没有运行 BGP。AS 63000 中运行的 IGP 协议是 RIP。

路由器 R3 在 AS 64000 中，并与路由器 R4 建立 EBGP 关系。R3 通告了 172.16.64.0/19 的路由；路由器 R7 在 AS 65000 中，并与路由器 R6 建立 EBGP 关系。R7 通告了 11.11.11.0/24 的路由。

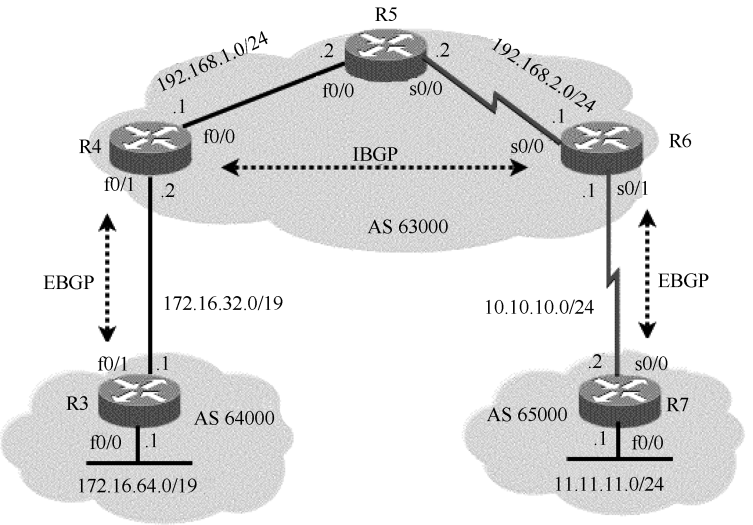

图 7-22　分析同步开启时使用的拓扑

（1）各路由器的相关配置

① 路由器 R3 的配置：

```
hostname "Router3"
!
interface FastEthernet0/0
 ip address 172.16.64.1 255.255.224.0
 duplex auto
 speed auto
!
interface FastEthernet0/1
 ip address 172.16.32.1 255.255.224.0
 duplex auto
 speed auto
```

```
!
router bgp 64000
 bgp log-neighbor-changes
 network 172.16.64.0 mask 255.255.224.0
 neighbor 172.16.32.2 remote-as 63000
```

② 路由器 R4 的配置：

```
hostname "router4"
!
interface FastEthernet0/0
 ip address 192.168.1.1 255.255.255.0
 duplex auto
 speed auto
!
interface FastEthernet0/1
 ip address 172.16.32.2 255.255.224.0
 duplex auto
 speed auto
!
router rip
 version 2
 network 172.16.0.0
 network 192.168.1.0
 no auto-summary
!
router bgp 63000
 bgp log-neighbor-changes
 neighbor 172.16.32.1 remote-as 64000
 neighbor 192.168.2.1 remote-as 63000
```

③ 路由器 R5 的配置：

```
hostname "router5"
!
interface FastEthernet0/0
 ip address 192.168.1.2 255.255.255.0
 duplex auto
 speed auto
!
interface Serial0/0
 ip address 192.168.2.2 255.255.255.0
!
router rip
 version 2
```

```
 network 192.168.1.0
 network 192.168.2.0
 no auto-summary
```

④ 路由器 R6 的配置：

```
hostname "router6"
!
interface Serial0/0
 ip address 192.168.2.1 255.255.255.0
!
interface Serial0/1
 ip address 10.10.10.1 255.255.255.0
!
router rip
 version 2
 network 10.0.0.0
 network 192.168.2.0
 no auto-summary
!
router bgp 63000
 bgp log-neighbor-changes
 neighbor 10.10.10.2 remote-as 65000
 neighbor 192.168.1.1 remote-as 63000
```

⑤ 路由器 R7 的配置：

```
hostname "router7"
!
interface FastEthernet0/0
 ip address 11.11.11.1 255.255.255.0
 duplex auto
 speed auto
!
interface Serial0/0
 ip address 10.10.10.2 255.255.255.0
!
router bgp 65000
 bgp log-neighbor-changes
 network 11.11.11.0 mask 255.255.255.0
 neighbor 10.10.10.1 remote-as 63000
```

这样配置路由器后，路由器 R3 通过 EBGP 把路由 172.16.64.0/19 通告给它的 EBGP 邻居 R4；路由器 R4 通过 IBGP 把该路由信息通告给它的 IBGP 邻居 R6，但不更改下一跳属性值。

根据 BGP 同步规则的要求，路由器 R6 要想使用并把该路由通告给它的 EBGP 邻居 R7，它必须满足同步，即该路由也必须存在于 IGP 数据库中。当前 AS 63000 中使用的 IGP 是 RIP，

也就是说，路由器 R6 也必须能够通过 RIP 学习到 172.16.64.0/19 的路由，否则，就不同步。

同理，路由器 R7 通过 EBGP 把路由 11.11.11.0/24 通告给它的 EBGP 邻居 R6；路由器 R6 通过 IBGP 把该路由信息通告给它的 IBGP 邻居 R4，但不更改下一跳属性值。路由器 R4 能够使用并向它的 EBGP 邻居 R3 通告该路由的前提条件是满足同步要求。

（2）路由器 R4 和 R6 的 BGP 库

① 路由器 R4 的 BGP 库：

```
router4#show ip bgp
BGP table version is 2, local router ID is 192.168.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
* i11.11.11.0/24   10.10.10.2              0     100      0      65000    i
*> 172.16.64.0/19  172.16.32.1             0              0      64000    i
router4#
```

输出显示，路由器 R4 虽然学习到了 11.11.11.0/24 的路由，但是该路由并非最佳路由，所以在该路由条目的状态标志中没有大于号（>）。数据库中也显示了下一跳地址是 10.10.10.2（R7 的接口地址）。从 R4 的路由表中可以看出下一跳是可达的。下面是 R4 的路由表。

② R4 的路由表：

```
router4#sh ip route
（代码省略）

     172.16.0.0/19 is subnetted, 2 subnets
C       172.16.32.0 is directly connected, FastEthernet0/1
B       172.16.64.0 [20/0] via 172.16.32.1, 00:16:33
     10.0.0.0/24 is subnetted, 1 subnets
R       10.10.10.0 [120/2] via 192.168.1.2, 00:00:13, FastEthernet0/0
C    192.168.1.0/24 is directly connected, FastEthernet0/0
R    192.168.2.0/24 [120/1] via 192.168.1.2, 00:00:13, FastEthernet0/0
router4#
```

输出显示，R4 的路由表中没有 11.11.11.0/24 的路由条目，但 10.10.10.0/24 的网络是可达的。因此，路由器 R4 之所以不使用 11.11.11.0/24 的路由是因为不同步，并非因为下一跳不可达。当然，R4 不会把该路由通告给它的 EBGP 邻居 R3。下面是 R3 的 BGP 库。

③ 路由器 R3 的 BGP 库：

```
Router3# sh ip bgp
BGP table version is 2, local router ID is 172.16.32.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? – incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
```

```
*> 172.16.64.0/19        0.0.0.0                  0         32768        i
Router3#
```

输出显示，R3 并没有得到 11.11.11.0/24 的路由。

同样的问题也会出现在路由器 R6 和 R7 上，下面是它们的 BGP 表和路由表。

④ 路由器 R6 的 BGP 库：

```
router6#sh ip bgp
BGP table version is 2, local router ID is 192.168.2.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
*> 11.11.11.0/24    10.10.10.2               0             0    65000    i
* i172.16.64.0/19   172.16.32.1              0     100     0    64000    i
router6#
```

输出显示，172.16.64.0/19 的路由不是最佳路由。

⑤ 路由器 R6 的路由表：

```
router6#sh ip route
（代码省略）

     172.16.0.0/19 is subnetted, 1 subnets
R       172.16.32.0 [120/2] via 192.168.2.2, 00:00:24, Serial0/0
     10.0.0.0/24 is subnetted, 1 subnets
C       10.10.10.0 is directly connected, Serial0/1
     11.0.0.0/24 is subnetted, 1 subnets
B       11.11.11.0 [20/0] via 10.10.10.2, 00:28:35
R    192.168.1.0/24 [120/1] via 192.168.2.2, 00:00:24, Serial0/0
C    192.168.2.0/24 is directly connected, Serial0/0
router6#
```

输出显示下一跳可达，没有 172.16.64.0/19 的 IGP 路由条目。

⑥ 路由器 R7 的 BGP 库：

```
router7#sh ip bgp
BGP table version is 2, local router ID is 10.10.10.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
*> 11.11.11.0/24    0.0.0.0                  0             32768    i
router7#
```

输出显示，R7 没有得到 172.16.64.0/19 的路由条目。

2. 同步开启且满足同步条件

为了满足同步条件，分别在路由器 R4 和路由器 R6 上设置静态路由。设置静态路由后的配置如下。

（1）路由器 R4 的相关配置

```
hostname "router4"
!
interface FastEthernet0/0
 ip address 192.168.1.1 255.255.255.0
 duplex auto
 speed auto
!
interface FastEthernet0/1
 ip address 172.16.32.2 255.255.224.0
 duplex auto
 speed auto
!
router rip
 version 2
 network 172.16.0.0
 network 192.168.1.0
 no auto-summary
!
router bgp 63000
 bgp log-neighbor-changes
 neighbor 172.16.32.1 remote-as 64000
 neighbor 192.168.2.1 remote-as 63000
!
ip route 11.11.11.0 255.255.255.0 192.168.1.2 200
```

增加了静态路由后，R4 的路由表如下所示。

（2）R4 的路由表

```
router4#sh ip route
（代码省略）

        172.16.0.0/19 is subnetted, 2 subnets
C           172.16.32.0 is directly connected, FastEthernet0/1
B           172.16.64.0 [20/0] via 172.16.32.1, 00:34:17
        10.0.0.0/24 is subnetted, 1 subnets
R           10.10.10.0 [120/2] via 192.168.1.2, 00:00:08, FastEthernet0/0
```

```
        11.0.0.0/24 is subnetted, 1 subnets
S       11.11.11.0 [200/0] via 192.168.1.2
C       192.168.1.0/24 is directly connected, FastEthernet0/0
R       192.168.2.0/24 [120/1] via 192.168.1.2, 00:00:08, FastEthernet0/0
router4#
```

这样，就满足了同步条件。下面是 R4 的 BGP 表。

（3）R4 的 BGP 表

```
router4#show ip bgp
BGP table version is 2, local router ID is 192.168.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
* > i11.11.11.0/24  10.10.10.2               0     100       0  65000    i
*> 172.16.64.0/19   172.16.32.1              0               0  64000    i
router4#
```

输出显示，11.11.11.0/24 的路由已经成为最佳路由。

再检查 R3 的 BGP 库。

（4）路由器 R3 的 BGP 库

```
Router3#sh ip bgp
BGP table version is 25, local router ID is 172.16.64.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
*> 11.11.11.0/24    172.16.32.2                              0  63000 65000  i
*> 172.16.64.0/19   0.0.0.0                  0              32768            i
Router3#
```

输出显示，路由器 R3 获得了 11.11.11.0/24 的路由。不用检查，路由表中也肯定出现了该路由条目。

同理，在路由器 R6 上也增加一条到达 172.16.64.0/19 的静态路由，设置静态路由后的 R6 的相关配置如下所述。

（5）R6 的配置

```
hostname "router6"
!
interface Serial0/0
 ip address 192.168.2.1 255.255.255.0
!
```

```
interface Serial0/1
 ip address 10.10.10.1 255.255.255.0
!
router rip
 version 2
 network 10.0.0.0
 network 192.168.2.0
 no auto-summary
!
router bgp 63000
 bgp log-neighbor-changes
 neighbor 10.10.10.2 remote-as 65000
 neighbor 192.168.1.1 remote-as 63000
!
ip route 172.16.64.0 255.255.224.0 Serial0/0
```

（6）R6 的路由表

```
router6# sh ip route
（代码省略）

     172.16.0.0/19 is subnetted, 2 subnets
R       172.16.32.0 [120/2] via 192.168.2.2, 00:00:10, Serial0/0
S       172.16.64.0 is directly connected, Serial0/0
     10.0.0.0/24 is subnetted, 1 subnets
C       10.10.10.0 is directly connected, Serial0/1
     11.0.0.0/24 is subnetted, 1 subnets
B       11.11.11.0 [20/0] via 10.10.10.2, 00:11:13
R    192.168.1.0/24 [120/1] via 192.168.2.2, 00:00:10, Serial0/0
C    192.168.2.0/24 is directly connected, Serial0/0
router6#
```

（7）R6 的 BGP 库

```
router6#sh ip bgp
BGP table version is 2, local router ID is 192.168.2.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
*> 11.11.11.0/24    10.10.10.2               0             0    65000    i
* > i172.16.64.0/19 172.16.32.1              0     100     0    64000    i
router6#
```

（8）路由器 R7 的 BGP 库

```
router7#sh ip bgp
BGP table version is 17, local router ID is 10.10.10.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
*> 11.11.11.0/24    0.0.0.0                  0              32768                 i
*> 172.16.64.0/19   10.10.10.1                              0    63000 64000      i
router7#
```

至此，问题好象解决了，其实不然。同步的问题虽然解决了，但是还存在其他问题。

从 R4 的 BGP 表显示，路由 11.11.11.0/24 的下一跳属性值是 10.10.10.2，那么，对于路由器 R4 来说该地址可达吗？该地址是否可达由 R4 的路由表来决定。再看 R4 的路由表。路由表显示，R4 使用自己的 f0/0 接口向网络 10.10.10.0/24 发送数据，所使用的下一跳地址是 192.168.1.2，即路由器 R5 的接口地址。这样看来，对于 R4 来说，10.10.10.2 的地址是可达的。也正是因为下一跳可达并且又满足了同步条件，所以 R4 在自己的 BGP 表中把 11.11.11.0/24 的路由标记为最佳。

如果路由器 R4 把目的地是网络 11.11.11.0/24 的数据发送给 R5，那么路由器 R5 有到达 11.11.11.0/24 的路由吗？下面是 R5 的路由表。

```
router5#sh ip route

（代码省略）
      172.16.0.0/19 is subnetted, 1 subnets
R        172.16.32.0 [120/1] via 192.168.1.1, 00:00:00, FastEthernet0/0
      10.0.0.0/24 is subnetted, 1 subnets
R        10.10.10.0 [120/1] via 192.168.2.1, 00:00:16, Serial0/0
C     192.168.1.0/24 is directly connected, FastEthernet0/0
C     192.168.2.0/24 is directly connected, Serial0/0
router5#
```

输出显示，路由器 R5 并没有到达 11.11.11.0/24 的路由，也没有到达 172.16.64.0/19 的路由。

因此，当 R5 收到目标地址是网络 11.11.11.0/24 和 172.16.64.0/19 的数据时，只能悄悄丢弃，这种现象称为“路由黑洞”。为了堵住这个路由“黑洞”，必须在 R5 上采取措施，比如设置静态路由，以保证整个网络是连通的。

下面是设置了静态路由后的 R5 的配置。

```
hostname "router5"
!
interface FastEthernet0/0
 ip address 192.168.1.2 255.255.255.0
 duplex auto
 speed auto
```

```
!
interface Serial0/0
 ip address 192.168.2.2 255.255.255.0
!
router rip
 version 2
 network 192.168.1.0
 network 192.168.2.0
 no auto-summary
!
ip route 11.0.0.0 255.0.0.0 Serial0/0
ip route 172.16.64.0 255.255.224.0 192.168.1.1
```

提示：

① 同步的路由前缀必须严格匹配，并且与 BGP 同步的 IGP 的管理距离不大于 200。

② 本例中使用静态路由的方法使 BGP 和 IGP 同步，以及在 R5 上使用静态路由保证不出现路由黑洞，只是为了分析 BGP 同步规则而采用的简单方法。此方法在实际的网络环境中不一定适用。实用的方法是路由再发布或者关闭同步。

3. 同步关闭

在 Cisco 路由器上，同步功能默认是开启的。关闭同步功能的命令语法如下：

```
router(config-router)# no synchronization
```

关闭同步功能后，只要下一跳可达，路由器即使用并向 EBGP 邻居通告来自 IBGP 邻居的路由信息。

因为在 AS 63000 中运行了 RIP 协议，保证了下一跳可达，所以，当在路由器 R4 和 R6 上关闭同步后，路由器 R4 立即向 R3 通告来自 R6 的 BGP 路由 11.11.11.0/24；路由器 R6 立即向 R7 通告来自 R4 的 BGP 路由 172.16.64.0/19。这样，路由器 R3 和 R7 就在它们的路由表中分别安装了相应的路由条目。

由于 R5 在转接路径上，所以，R5 也必须有 11.11.11.0/24 的路由和 172.16.64.0/19 的路由，否则会出现“路由黑洞”。因此，R5 上也需要运行 BGP，让它从 R4 和 R6 处获得路由，即只有在转接路径上的所有路由器都运行了 BGP 时，关闭同步才是安全的。关闭同步后的拓扑如图 7-23 所示。

路由器 R3 和 R7 的配置不变，下面是路由器 R4、R5、R6 的配置。

（1）路由器 R4 的相关配置

```
hostname "router4"
!
interface FastEthernet0/0
 ip address 192.168.1.1 255.255.255.0
 duplex auto
```

```
 speed auto
!
interface FastEthernet0/1
 ip address 172.16.32.2 255.255.224.0
 duplex auto
 speed auto
!
router rip
 version 2
 network 172.16.0.0
 network 192.168.1.0
 no auto-summary
!
router bgp 63000
 no synchronization
 bgp log-neighbor-changes
 neighbor 172.16.32.1 remote-as 64000
 neighbor 192.168.1.2 remote-as 63000
 neighbor 192.168.2.1 remote-as 63000
```

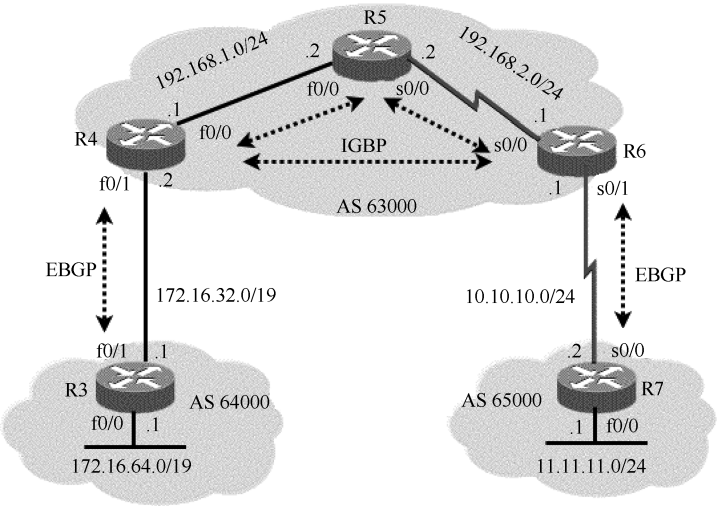

图 7-23　分析同步关闭时使用的拓扑

（2）路由器 R5 的配置

```
hostname "router5"
!
interface FastEthernet0/0
 ip address 192.168.1.2 255.255.255.0
 duplex auto
 speed auto
```

```
!
interface Serial0/0
 ip address 192.168.2.2 255.255.255.0
!
router rip
 version 2
 network 192.168.1.0
 network 192.168.2.0
 no auto-summary
!
router bgp 63000
 no synchronization
 bgp log-neighbor-changes
 neighbor 192.168.1.1 remote-as 63000
 neighbor 192.168.2.1 remote-as 63000
```

（3）路由器 R6 的配置

```
hostname "router6"
!
interface Serial0/0
 ip address 192.168.2.1 255.255.255.0
!
interface Serial0/1
 ip address 10.10.10.1 255.255.255.0
!
router rip
 version 2
 network 10.0.0.0
 network 192.168.2.0
 no auto-summary
!
router bgp 63000
 no synchronization
 bgp log-neighbor-changes
 neighbor 10.10.10.2 remote-as 65000
 neighbor 192.168.1.1 remote-as 63000
 neighbor 192.168.2.2 remote-as 63000
```

下面是这 3 台路由器的 BGP 库及路由表。

（4）路由器 R4 的 BGP 库

```
router4# sh ip bgp
BGP table version is 15, local router ID is 192.168.1.1
```

```
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*>i11.11.11.0/24    10.10.10.2               0    100      0 65000 i
*> 172.16.64.0/19   172.16.32.1              0             0 64000 i
router4#
```

（5）查看路由器 R4 的一条具体路由信息

```
router4#sh ip bgp 172.16.64.0
1. BGP routing table entry for 172.16.64.0/19, version 3
2. Paths: (1 available, best #1, table Default-IP-Routing-Table)
3. Advertised to non peer-group peers:
4. 192.168.1.2 192.168.2.1
5. 64000
6. 172.16.32.1 from 172.16.32.1 (172.16.64.1)
7. Origin IGP, metric 0, localpref 100, valid, external, best
router4#
```

该命令输出的各项含义如下所述。

第 1 行显示，当前的 BGP 路由表更新版本号为 3，而且是路由条目 172.16.64.0/19 的信息。

第 2 行显示的是路径信息。有 1 条可用路径，最佳路径是 1 号路径（实际上就是这条可用的路径）。最佳路径已经安装到路由表中（Table Default-IP-Routing-Table）。

第 3 行显示，该最佳路径通告给了非对等体组对等体，它们是 192.168.1.2 和 192.168.2.1。

第 5 行显示的是这条路由的 AS 路径属性值。

第 6 行显示了通告该路由信息的路由器的接口地址及路由器 ID（172.16.64.1）。

第 7 行显示了该路由的其他属性：起源于 IGP；度量为 0；本地优先为 100；是有效路由；是外部路由；是最佳路由。

（6）路由器 R4 的路由表

```
router4#sh ip route
（代码省略）

     172.16.0.0/19 is subnetted, 2 subnets
C       172.16.32.0 is directly connected, FastEthernet0/1
B       172.16.64.0 [20/0] via 172.16.32.1, 00:51:28
     10.0.0.0/24 is subnetted, 1 subnets
R       10.10.10.0 [120/2] via 192.168.1.2, 00:00:10, FastEthernet0/0
     11.0.0.0/24 is subnetted, 1 subnets
B       11.11.11.0 [200/0] via 10.10.10.2, 00:10:30
C    192.168.1.0/24 is directly connected, FastEthernet0/0
R    192.168.2.0/24 [120/1] via 192.168.1.2, 00:00:10, FastEthernet0/0
router4#
```

（7）路由器 R5 的 BGP 库

```
router5#sh ip bgp
BGP table version is 3, local router ID is 192.168.2.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*>i11.11.11.0/24    10.10.10.2               0    100      0 65000 i
*>i172.16.64.0/19   172.16.32.1              0    100      0 64000 i
router5#
```

（8）路由器 R5 的路由表

```
router5#sh ip route
（代码省略）

172.16.0.0/19 is subnetted, 2 subnets
R       172.16.32.0 [120/1] via 192.168.1.1, 00:00:02, FastEthernet0/0
B       172.16.64.0 [200/0] via 172.16.32.1, 00:00:11
     10.0.0.0/24 is subnetted, 1 subnets
R       10.10.10.0 [120/1] via 192.168.2.1, 00:00:04, Serial0/0
     11.0.0.0/24 is subnetted, 1 subnets
B       11.11.11.0 [200/0] via 10.10.10.2, 00:00:11
C    192.168.1.0/24 is directly connected, FastEthernet0/0
C    192.168.2.0/24 is directly connected, Serial0/0
router5#
```

（9）路由器 R6 的 BGP 库

```
router6#sh ip bgp
BGP table version is 29, local router ID is 192.168.2.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 11.11.11.0/24    10.10.10.2               0             0 65000 i
*>i172.16.64.0/19   172.16.32.1              0    100      0 64000 i
router6#
```

（10）路由器 R6 的路由表

```
router6# sh ip route
（代码省略）
```

```
172.16.0.0/19 is subnetted, 2 subnets
R       172.16.32.0 [120/2] via 192.168.2.2, 00:00:09, Serial0/0
B       172.16.64.0 [200/0] via 172.16.32.1, 00:01:22
    10.0.0.0/24 is subnetted, 1 subnets
C       10.10.10.0 is directly connected, Serial0/1
    11.0.0.0/24 is subnetted, 1 subnets
B       11.11.11.0 [20/0] via 10.10.10.2, 00:32:29
R    192.168.1.0/24 [120/1] via 192.168.2.2, 00:00:09, Serial0/0
C    192.168.2.0/24 is directly connected, Serial0/0
router6#
```

根据这 3 台路由器的 BGP 库和路由表，可以断定路由器 R3 和 R7 的数据库和路由表是正确的，这里不再列出。

提示：只有在转接路径上的所有路由器都运行了 BGP 时，关闭同步才是安全的，否则可能出现路由黑洞。

7.10.7　配置路由反射器

在图 7-23 所示的拓扑中，路由器 R4、R5 和 R6 之间建立了全互连的 IBGP 关系，目的是绕过由 BGP Split Horizon 规则带来的限制。因为 R5 不会把来自 R4 的 BGP 路由传递给 R6，也不会把来自 R6 的 BGP 路由传递给 R4。不足之处是需要更多的 BGP 会话。

可以使用路由反射器技术来解决 TCP 会话过多的问题。在设置路由反射器时，只需在充当路由反射器的路由器上声明哪个邻居是它的客户端即可，在客户端路由器上不需要任何配置。

① 配置路由反射器的命令语法如下：

```
router(config-router)# neighbor ip-address route-reflector-client
```

例如，在图 7-24 所示的拓扑中，路由器 R5 作为路由反射器是恰当的。

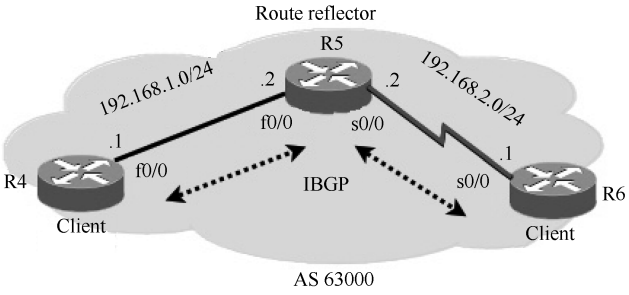

图 7-24　路由反射器的配置

路由器 R5 的配置如下（只列出相关配置）：

```
router bgp 63000
 no synchronization
 bgp log-neighbor-changes
```

```
 neighbor 192.168.1.1 remote-as 63000
 neighbor 192.168.1.1 route-reflector-client
 neighbor 192.168.2.1 remote-as 63000
 neighbor 192.168.2.1 route-reflector-client
```

配置了路由反射器后，路由器 R4 和 R6 之间就可以不建立 IBGP 邻接关系了。

② 使用 show ip bgp neighbors 命令可以检查邻居是路由反射器或者是客户端。例如：

```
router5#sh ip bgp neighbors
BGP neighbor is 192.168.1.1,  remote AS 63000, internal link
  BGP version 4, remote router ID 192.168.1.1
  BGP state = Established, up for 00:06:20
  Last read 00:00:20, hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received(old & new)
    Address family IPv4 Unicast: advertised and received
  Received 38 messages, 0 notifications, 0 in queue
  Sent 35 messages, 0 notifications, 0 in queue
  Route refresh request: received 0, sent 0
  Default minimum time between advertisement runs is 5 seconds

 For address family: IPv4 Unicast
  BGP table version 11, neighbor version 11
  Index 1, Offset 0, Mask 0x2
  Route-Reflector Client
……
```

输出显示，邻居是客户端；也说明本台路由器是路由反射器。

7.10.8　通告聚合路由

为了使路由表中安装较少的路由条目，提高路由器查询路由表的速度，把多条前缀较长的路由条目聚合（Aggregate）成前缀更短的一条路由是理想的方法。这与 OSPF 中的路由总结的含义相同。

在 BGP 路由器上执行路由聚合的命令语法如下：

```
router(config-router)# aggregate-address network-address mask [as-set] [summary-only] [suppress-map route-map-name] [advertise-map route-map-name] [attribute-map route-map-name]
```

- *address*：聚合后的路由前缀。
- *mask*：聚合路由掩码。
- ［as-set］：默认地，聚合路由被认为是始发自执行聚合的 AS，在 AS 路径中只列出执行聚合的 AS 号。如果使用该可选参数，在聚合路由的 AS 路径属性中将携带被聚合路由的 AS 号，只不过这些 AS 号的排列是无序的。重要的是，使用该参数后，聚合路由将继承被聚合路由的属性。
- ［summary-only］：只通告聚合路由，否则将一同通告聚合路由和被聚合的那些具体

路由条目。

- ［suppress-map *route-map-name*］：使用该选项调用一个路由图，可以选择性地抑制某些具体路由条目（被聚合的路由）而通告另外一些具体路由条目。
- ［advertise-map *route-map-name*］：使用该选项调用一个路由图，可以定义使用那些具体路由的属性值作为聚合路由的属性值。
- ［attribute-map *route-map-name*］：使用该选项调用一个路由图，可以更改聚合路由的属性值。
- *route-map-name*：被调用的路由图的名字。有关路由图的内容参见第 8 章。

提示：该命令聚合的是那些已经存在于 BGP 表中的路由条目，它与使用 network 命令通告一条总结的路由是不同的。network 命令通告的总结路由应该是 IGP 的总结路由，这些路由应该存在于路由表中。

例如，在图 7-25 所示的网络中，路由器 R2 把 172.16.96.0/19、172.16.128.0/19、172.16.160.0/19、172.16.192.0/19 直连网络用 network 命令通告到 BGP，然后对这些 19 位长的前缀路由进行聚合，并把前缀聚合到 16 位长。

（1）路由器 R2 的配置

```
hostname router2
!
interface Loopback0
 ip address 172.16.96.1 255.255.224.0
!
interface Loopback1
 ip address 172.16.128.1 255.255.224.0
!
interface Loopback2
 ip address 172.16.160.1 255.255.224.0
!
interface Loopback3
 ip address 172.16.192.1 255.255.224.0
!
interface FastEthernet1/0
 ip address 172.16.64.2 255.255.224.0
 duplex auto
 speed auto
!
router bgp 64000
 network 172.16.96.0 mask 255.255.224.0
 network 172.16.128.0 mask 255.255.224.0
 network 172.16.160.0 mask 255.255.224.0
```

```
network 172.16.192.0 mask 255.255.224.0
aggregate-address 172.16.0.0 255.255.0.0 as-set summary-only
neighbor 172.16.64.1 remote-as 64000
```

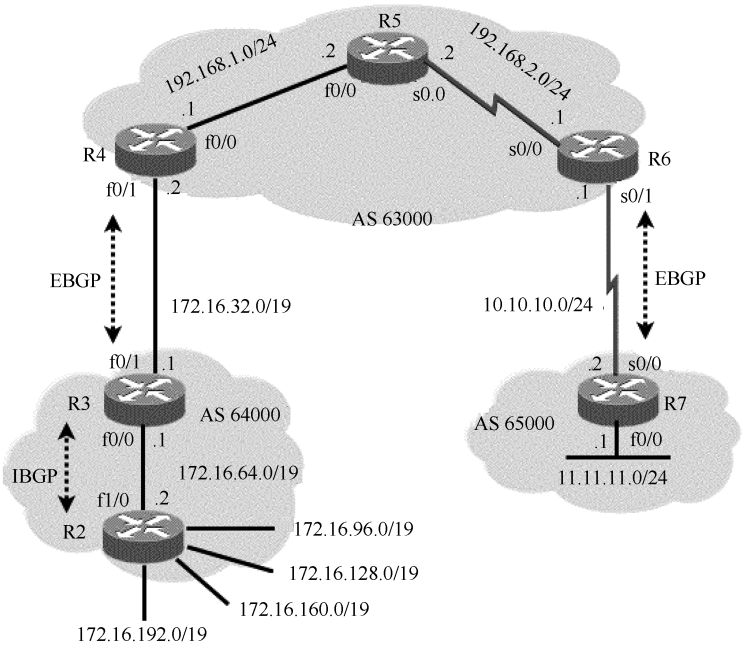

图 7-25　路由聚合使用的拓扑

（2）检查路由器运行情况的输出

① 路由器 R2 的 BGP 表：

```
router2# sh ip bgp
BGP table version is 16, local router ID is 172.16.192.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
* i11.11.11.0/24   172.16.32.2               100       0  63000 65000  i
*> 172.16.0.0       0.0.0.0                  100   32768  i
s>i172.16.64.0/19  172.16.64.1              0  100       0  i
s> 172.16.96.0/19  0.0.0.0                  0         32768  i
s> 172.16.128.0/19 0.0.0.0                  0         32768  i
s> 172.16.160.0/19 0.0.0.0                  0         32768  i
s> 172.16.192.0/19 0.0.0.0                  0         32768  i
router2#
```

输出显示，由路由器 R3 通告过来的 172.16.64.0/19 也被聚合了。

② 路由器 R2 的路由表：

```
router2#    sh ip route
```

```
（代码省略）

172.16.0.0/16 is variably subnetted, 6 subnets, 2 masks
C        172.16.160.0/19 is directly connected, Loopback2
C        172.16.128.0/19 is directly connected, Loopback1
C        172.16.192.0/19 is directly connected, Loopback3
B        172.16.0.0/16 [200/0] via 0.0.0.0, 00:04:10, Null0
C        172.16.96.0/19 is directly connected, Loopback0
C        172.16.64.0/19 is directly connected, FastEthernet1/0
router2#
```

为了避免路由环路的出现，路由器自动在路由表中增加一条下一跳为 null0 的聚合路由条目。

③ 路由器 R3 的 BGP 表：

```
Router3#sh ip bgp
BGP table version is 52, local router ID is 172.16.64.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
*> 11.11.11.0/24    172.16.32.2                              0  63000 65000  i
*>i172.16.0.0       172.16.64.2                    100       0               i
*> 172.16.64.0/19   0.0.0.0                  0               32768           i
Router3#
```

由于在路由器 R2 的聚合路由命令中使用了 summary-only 参数，所以，路由器 R3 只收到了聚合路由 172.16.0.0/16。

④ 路由器 R4 的 BGP 表：

```
router4#sh ip bgp
BGP table version is 40, local router ID is 192.168.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
*>i11.11.11.0/24    10.10.10.2               0     100       0  65000  i
*> 172.16.0.0       172.16.32.1                              0  64000  i
*> 172.16.64.0/19   172.16.32.1              0               0  64000  i
router4#
```

输出显示，R4 收到的也是聚合路由。

⑤ 路由器 R4 的路由表：

```
router4#sh ip route
 （代码省略）

172.16.0.0/16 is variably subnetted, 3 subnets, 2 masks
```

```
C        172.16.32.0/19 is directly connected, FastEthernet0/1
B        172.16.0.0/16 [20/0] via 172.16.32.1, 00:01:49
B        172.16.64.0/19 [20/0] via 172.16.32.1, 00:01:49
      10.0.0.0/24 is subnetted, 1 subnets
R        10.10.10.0 [120/2] via 192.168.1.2, 00:00:23, FastEthernet0/0
      11.0.0.0/24 is subnetted, 1 subnets
B        11.11.11.0 [200/0] via 10.10.10.2, 00:01:49
C     192.168.1.0/24 is directly connected, FastEthernet0/0
R     192.168.2.0/24 [120/1] via 192.168.1.2, 00:00:23, FastEthernet0/0
router4#
```

7.10.9　配置对等体组

1. 基本概念

共享相同外出策略（Outbound Policies）的 BGP 对等体可以被组织在一起称为对等体组（Peer Group）。定义对等体组的主要优点是降低路由更新使用的系统资源（如 CPU 和内存），简化配置。路由器只需检查一次路由表，就可以把更新复制给对等体组中的每一个成员，而不需要为每一个对等体单独进行更新。根据对等体组成员的数量、路由前缀的数量和被更新的前缀数量，对等体组能够明显地降低更新负载量。对等体组也允许把相同的外出策略应用到对等体组，而不需要一个一个地应用到单个对等体。对等体组中的成员继承对等体组的策略和配置。

定义对等体组必须满足以下条件：

- 所有组成员必须是具有相同外出更新策略的对等体。
- 组成员可以有独立的入（Inbound）更新策略。
- 对等体组可以是内部对等体组（所有成员必须是 IBGP 邻接）或者外部对等体组（所有成员必须是 EBGP 邻接）。

2. 配置命令

配置对等体组的过程可以分为以下 3 步。

- 生成对等体组；
- 向对等体组指派成员；
- 应用策略（参见第 8 章）。

① 生成对等体组的命令语法如下：

```
router(config-router)# neighbor peer-group-name peer-group
```

- *peer-group-name*：对等体组的名字。

② 向对等体组指派成员的命令语法如下：

```
router(config-router)# neighbor ip-address peer-group peer-group-name
```

- *ip-address*：对等体成员的地址；
- *peer-group-name*：将要加入的对等体组名字。

默认地，所有组成员继承对等体组的所有配置，组成员也可以单独配置一些不影响外出更

新策略的设置选项，以覆盖继承来的选项。

组成员总是继承以下配置选项：

- remote-as；
- version；
- update-source；
- out-route-map （控制外出数据的路由图）；
- out-filter-list （控制外出数据的过滤列表）；
- out-distribute-list （控制外出数据的发布列表）；
- minimum-advertisement-interval；
- next-hop-self。

组成员也继承对对等体组做的修改。如果对等体组没有配置 remote-as 选项，组成员可以使用“**neighbor** {*ip-address* | *peer-group-name*} **remote-as**” 命令单独配置。

3. 对等体组配置示例

（1）IBGP 对等体组示例

从“router bgp 100”和“neighbor internal remote-as 100”可以看出，该示例中的组成员都在 AS 100 中，对等体组的名字是“internal”。所有的组成员都使用 loopback0 接口建立 BGP 会话。对等体组应用了名字为“set-med”的路由图和“filter-list 1”的外出路由策略。另外，成员 172.16.232.55 还应用了一个入路由更新策略。

```
router bgp 100
neighbor internal peer-group
neighbor internal remote-as 100
neighbor internal update-source loopback 0
neighbor internal route-map set-med out
neighbor internal filter-list 1 out
neighbor 172.16.232.53 peer-group internal
neighbor 172.16.232.54 peer-group internal
neighbor 172.16.232.55 peer-group internal
neighbor 172.16.232.55 filter-list 3 in
```

（2）EBGP 对等体组示例

```
router bgp 100
neighbor external-peers peer-group
neighbor external-peers route-map set-metric out
neighbor external-peers filter-list 99 out
neighbor external-peers filter-list 101 in
neighbor 172.16.232.90 remote-as 200
neighbor 172.16.232.90 peer-group external-peers
neighbor 172.16.232.100 remote-as 300
neighbor 172.16.232.100 peer-group external-peers
```

```
neighbor 172.16.232.110 remote-as 400
neighbor 172.16.232.110 peer-group external-peers
neighbor 172.16.232.110 filter-list 400 in
```

配置显示这是一个 EBGP 对等体组，组成员分别位于 AS 200、AS 300、AS 400 中。对等体组应用的外出路由更新策略是名字是 set-metric 的路由图和 filter-list 99。同时，该对等体组还应用了一个入路由策略——filter-list 101。另外，路由器 172.16.232.110 还单独应用了一个入路由策略。

提示：在实现对等体组之前，你应该查阅特定 IOS 版本对其有何限制。

7.10.10　应用团体属性

在 BGP 路由器上使用团体属性，需要做以下工作：

- 采用新的表示格式（可选）；
- 配置路由器发送团体属性；
- 配置团体列表（可选）；
- 使用路由图设置团体属性值（Route-map 的配置详见第 8 章）。

① 在 CISCO 路由器上，默认的团体标志号表示格式是一个十进制数，RFC1997 定义的新的团体标志号格式是 AA:NN。AA 代表自治系统号，NN 代表团体号，各占 2 字节。使用如下命令让路由器使用新格式：

```
Router(config)# ip bgp-community new-format
```

② 发送团体属性：

默认地，团体属性在外出的路由更新中被剥去，不发送到任何对等体。如果需要让路由器向它的对等体发送团体属性，使用如下命令开启：

```
Router(config-router)# neighbor {ip-address | peer-group-name} send-community [both | standard | extended]
```

- *ip-address*：对等体的地址。
- *peer-group-name*：对等体组的名字。
- **both**：标准和扩展团体属性。
- **standard**：仅标准团体属性。
- **extended**：仅扩展团体属性。

③ 配置团体列表：

团体列表（Community-list）的功能类似于 ACL，不同的是团体列表中的条件是团体标志号而不是 IP 地址。路由图可以调用团体列表，根据团体列表中的定义为路由设置某些属性。配置团体列表的命令语法如下：

```
router(config)# ip community-list {standard-number | standard list-name {deny | permit}
[community-number] [AA:NN] [internet] [local-AS] [no-advertise] [no-export]} | {expanded-number | expanded
list-name {deny | permit} regexp}
```

- *standard-number*：标准团体列表号，1～99。
- **standard** *list-name*：标准团体列表名字。
- **deny**：匹配团体属性值的的路由被禁止。

- **permit**：匹配团体属性值的的路由被允许。
- *community-number*] [*AA*:*NN*]：团体标志号。
- *expanded-number*：扩展团体列表号，100～500。
- **expanded** *list-name*：扩展团体列表名字。
- *regexp*：扩展团体列表中使用的通用表达式。

标准团体列表用来定义公认的团体及一些一般的团体。标准团体列表中最多只能罗列 16 个团体号。

扩展团体列表可以使用通用逻辑表达式定义团体号，具有更大的灵活性和匹配能力。有关通用表达式的使用方法，已超出本书范围，请读者自行查阅 Cisco 文档。

下面是团体列表的示例。

① 同时携带团体属性值为 65534:40 和 65412:60 的路由被禁止。

```
Router(config)# ip community-list 2 deny 65534:40 65412:60
```

② 所有本自治系统的路由或者来自 AS 40000 团体号为 20 的路由被允许。

```
Router(config)# ip community-list standard RED permit local-AS
Router(config)# ip community-list standard RED permit 40000:20
```

7.10.11　其他命令

（1）修改本地优先属性值

Cisco 路由器的默认本地优先属性值为 100，可以使用命令修改该值，命令语法如下：

```
Router(config-router)# bgp default local-preference value
```

- *value*：本地属性值，0～4 294 967 295。

（2）修改权重属性值

为来自不同的邻居的路由指派权重值使用如下命令：

```
router(config-router)# neighbor {ip-address | peer-group-name} weight weight
```

- *ip-address*：邻居地址。
- *peer-group-name*：对等体组名字。
- *weight*：权重值，0～65 535。

（3）快速重置 EBGP 会话

如果与 EBGP 对等体的直连链路 Down，使用如下命令让路由器立即重置会话。

```
router(config-router)# bgp fast-external-fallover
```

（4）设置路由器 ID

```
router(config-router)# bgp router-id ip-address
```

- *ip-address*：路由器 ID。

（5）重置及软重置 BGP 会话

```
router(config)# clear ip bgp {* | address | peer-group-name} [soft [in | out]]
```

- *：重置所有 BGP 会话。
- *address*：邻居地址。
- *peer-group-name*：对等体组名字。
- soft：软件意义上的重置不重置 TCP 会话，只进行数据库更新。
- in：入方向上的更新。
- out：出方向上的更新。

在以下情况下需要重置或软重置 BGP：

- 与 BGP 相关的访问列表的添加或更改；
- 与 BGP 相关的发布列表的更改（发布列表在第 8 章解释）；
- 与 BGP 相关的路由图的更改。

（6）终止与邻居的 BGP 会话

```
router(config-router)# neighbor {ip-address | peer-group-name} shutdown
```

（7）显示团体列表允许的路由条目

```
router# show ip bgp community-list community-list-number [exact]
```

- [exact]：只列出精确匹配的路由条目。

（8）修改计时器

```
router(config-router)# timers bgp keepalive holdtime
```

- *keepalive*：keepalive 发送间隔，默认为 60 秒。
- *holdtime*：保持时间，应为 *keepalive* 时间的 3 倍。

7.11　多　宿　主

BGP 协议是一个与 Internet 建立冗余连接的关键协议之一。当用户的 AS 与两个以上不同的 ISP 相连时称为多宿主（Multihoming），如图 7-26 和图 7-27 所示。

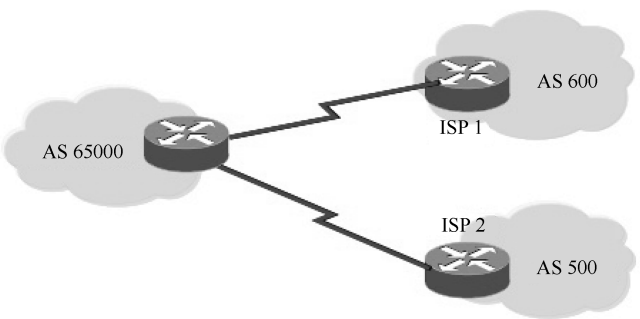

图 7-26　多宿主连接方式

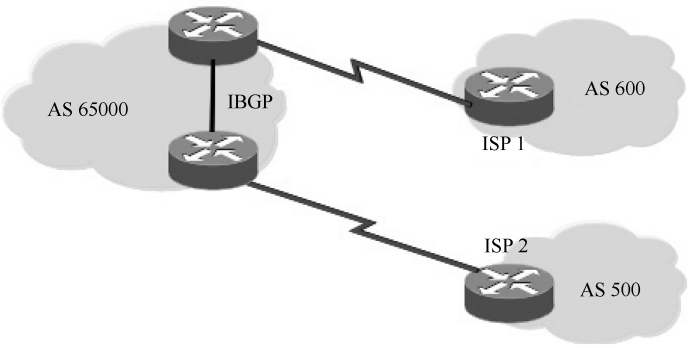

图 7-27　多宿主连接类型

多宿主可以提供冗余的链路，当与一个 ISP 的链路出现故障时，用户可以使用与另一个 ISP 连接的链路传送数据。多宿主还可以优化网络提高性能。它可以选择为其提供最佳路径的信息的 ISP 转发数据。多宿主的不足之处是用户的 AS 面临着成为 ISP 之间流量的转递 AS，如果这种情况发生，来自 Internet 的流量将把用户路由器的系统资源，如 CPU、内存等耗费怠尽。因此，实现时需要仔细规划，避免成为 ISP 的转递 AS。

在多宿主环境中，对路由信息的处理有以下 3 种情况：

① 所有 ISP 发送所有路由给用户 AS，用户向 ISP 发送所有路由。

这种情况下，用户的路由器需要有足够的系统资源，因为 ISP 传递来的路由信息可能相当多。

用户自治系统内的路由器将根据自治系统内运行的 IGP 把内部数据送达运行 BGP 的路由器，BGP 路由器通常根据 AS 路径属性值选择 ISP 把数据发送到外部网络，也可以使用权重属性和本地优先属性选择 ISP。进入用户 AS 的数据所采纳的路径由 ISP 决定。

② 所有 ISP 只发默认路由给用户 AS，用户向 ISP 发送所有路由。

这种情况对用户路由器要求最低，因为它只需要处理默认路由即可。用户自治系统内的路由器将根据自治系统内运行的 IGP 把内部数据送达运行 BGP 的路由器，BGP 路由器使用来自 ISP 的默认路由把数据送达自治系统外。进入用户 AS 的数据所采纳的路径由 ISP 决定。

③ 所有 ISP 向用户 AS 发送某些特定路由和默认路由。用户向 ISP 发送所有路由。

用户自治系统内的路由器将根据自治系统内运行的 IGP 把内部数据送达运行 BGP 的路由器，由 BGP 路由器决定如何把数据送达自治系统外。对于那些到达特定网络的数据，BGP 路由器查看哪个 ISP 通告的特定路由具有选择最短的 AS 路径，然后使用该 ISP 作为外出数据的出口（也可以使用权重和本地优先属性）。对于没有特定路由的数据 BGP 路由器，使用默认路由将其发送到自治系统外。进入用户 AS 的数据所采纳的路径由 ISP 决定。

7.12　BGP 案例

有 AS 65000、AS 65500 和 AS 645003 个自治系统互连，拓扑结构如图 7-28 所示，每个自治系统内的路由器只需要运行 IGP，保证本自治系统内的网络可达。自治系统 AS 65000 和 AS 65500 运行的 IGP 是 IS-IS；自治系统 AS 64500 运行的 IGP 是 RIP。自治系统之间的可达性由 BGP 协议保证，因此，在自治系统边界上的路由器既运行 IGP 又运行 BGP，并负责把本自治

系统内的路由通告给其他 AS。为了保证各自治系统内运行 IGP 的路由器能够有路由转发到达其他自治系统的数据，边界路由器向 IGP 中注入了默认路由。

每个自治系统内部还有其他路由器及其他网络，本例中没有列出。

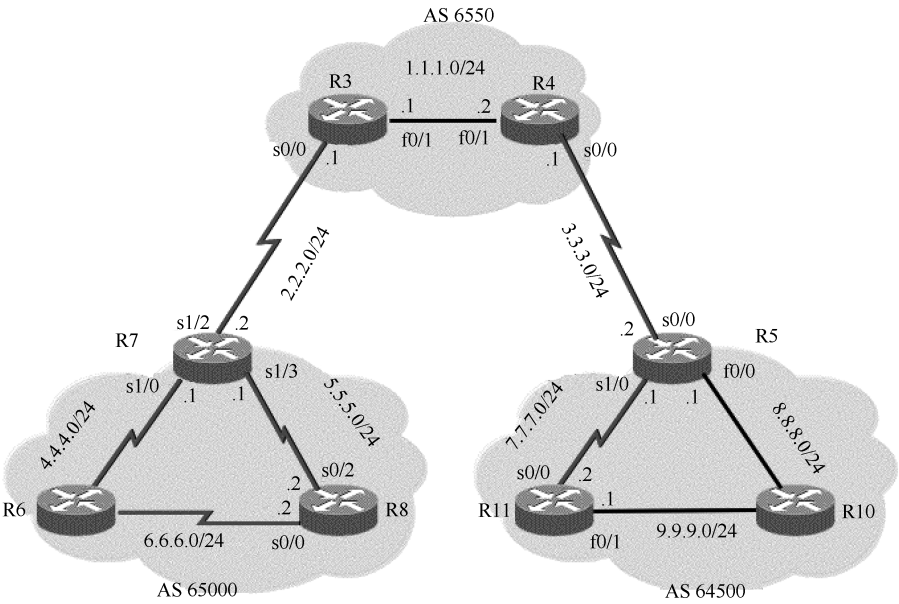

图 7-28　BGP 案例使用的拓扑

下面是部分路由器的相关配置、BGP 表和路由表。路由器 R6 的配置和路由器 R8 的相似；路由器 R10 的配置和路由器 R11 的相似。

（1）路由器 R3 的配置

```
hostname "Router3"
!
interface Serial0/0
 ip address 2.2.2.1 255.255.255.0
 no fair-queue
 clockrate 64000
!
interface FastEthernet0/1
 ip address 1.1.1.1 255.255.255.0
 ip router isis
 duplex auto
 speed auto
!
router isis
 net 49.0001.0000.0000.0016.00
!
router bgp 65500
```

```
 no synchronization
 bgp log-neighbor-changes
 neighbor 1.1.1.2 remote-as 65500
 neighbor 1.1.1.2 next-hop-self
 neighbor 2.2.2.2 remote-as 65000
```

（2）路由器 R3 的 BGP 表

```
Router3#sh ip bgp
BGP table version is 34, local router ID is 2.2.2.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
*>i1.1.1.0/24       1.1.1.2                  0     100       0         i
*> 2.2.2.0/24       2.2.2.2                  0               0   65000 i
*>i3.3.3.0/24       1.1.1.2                  0     100       0   64500 i
*> 4.4.4.0/24       2.2.2.2                  0               0   65000 i
*> 5.5.5.0/24       2.2.2.2                  0               0   65000 i
*> 6.6.6.0/24       2.2.2.2                 20               0   65000 i
*>i7.7.7.0/24       1.1.1.2                  0     100       0   64500 i
*>i8.8.8.0/24       1.1.1.2                  0     100       0   64500 i
*>i9.0.0.0          1.1.1.2                  1     100       0   64500 i
Router3#
```

（3）路由器 R3 的路由表

```
Router3#sh ip route
（代码省略）
     1.0.0.0/24 is subnetted, 1 subnets
C       1.1.1.0 is directly connected, FastEthernet0/1
     2.0.0.0/24 is subnetted, 1 subnets
C       2.2.2.0 is directly connected, Serial0/0
     3.0.0.0/24 is subnetted, 1 subnets
B       3.3.3.0 [200/0] via 1.1.1.2, 00:22:57
     4.0.0.0/24 is subnetted, 1 subnets
B       4.4.4.0 [20/0] via 2.2.2.2, 05:44:57
     5.0.0.0/24 is subnetted, 1 subnets
B       5.5.5.0 [20/0] via 2.2.2.2, 05:43:57
     6.0.0.0/24 is subnetted, 1 subnets
B       6.6.6.0 [20/20] via 2.2.2.2, 05:43:57
     7.0.0.0/24 is subnetted, 1 subnets
B       7.7.7.0 [200/0] via 1.1.1.2, 00:22:57
```

```
        8.0.0.0/24 is subnetted, 1 subnets
B          8.8.8.0 [200/0] via 1.1.1.2, 00:22:57
B       9.0.0.0/8 [200/1] via 1.1.1.2, 00:22:53
Router3#
```

（4）路由器 R4 的配置

```
hostname "R4"
!
interface Serial0/0
 ip address 3.3.3.1 255.255.255.0
 no fair-queue
 clockrate 64000
!
interface FastEthernet0/1
 ip address 1.1.1.2 255.255.255.0
 ip router isis
 duplex auto
 speed auto
!
router isis
 net 49.0001.0000.0000.0123.00
!
router bgp 65500
 no synchronization
 bgp log-neighbor-changes
 network 1.1.1.0 mask 255.255.255.0
 neighbor 1.1.1.1 remote-as 65500
 neighbor 1.1.1.1 next-hop-self
 neighbor 3.3.3.2 remote-as 64500
```

（5）路由器 R4 的 BGP 表

```
R4#sh ip bgp
BGP table version is 362, local router ID is 192.168.55.120
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric  LocPrf  Weight  Path
*> 1.1.1.0/24       0.0.0.0                  0          32768         i
*>i2.2.2.0/24       1.1.1.1                  0     100      0    65000 i
*> 3.3.3.0/24       3.3.3.2                  0              0    64500 i
*>i4.4.4.0/24       1.1.1.1                  0     100      0    65000 i
```

```
*>i5.5.5.0/24        1.1.1.1                  0    100      0    65000    i
*>i6.6.6.0/24        1.1.1.1                 20    100      0    65000    i
*> 7.7.7.0/24        3.3.3.2                  0             0    64500    i
*> 8.8.8.0/24        3.3.3.2                  0             0    64500    i
*> 9.0.0.0           3.3.3.2                  1             0    64500    i
R4#
```

（6）路由器 R4 的路由表

```
R4#sh ip route
（代码省略）

     1.0.0.0/24 is subnetted, 1 subnets
C       1.1.1.0 is directly connected, FastEthernet0/1
     2.0.0.0/24 is subnetted, 1 subnets
B       2.2.2.0 [200/0] via 1.1.1.1, 00:47:33
     3.0.0.0/24 is subnetted, 1 subnets
C       3.3.3.0 is directly connected, Serial0/0
     4.0.0.0/24 is subnetted, 1 subnets
B       4.4.4.0 [200/0] via 1.1.1.1, 00:47:33
     5.0.0.0/24 is subnetted, 1 subnets
B       5.5.5.0 [200/0] via 1.1.1.1, 00:47:33
     6.0.0.0/24 is subnetted, 1 subnets
B       6.6.6.0 [200/20] via 1.1.1.1, 00:47:33
     7.0.0.0/24 is subnetted, 1 subnets
B       7.7.7.0 [20/0] via 3.3.3.2, 00:01:26
     8.0.0.0/24 is subnetted, 1 subnets
B       8.8.8.0 [20/0] via 3.3.3.2, 00:01:26
C    192.168.55.0/24 is directly connected, FastEthernet0/0
B    9.0.0.0/8 [20/1] via 3.3.3.2, 00:01:27
R4#
```

（7）路由器 R5 的配置

```
hostname "R5"
!
interface FastEthernet0/0
 ip address 8.8.8.1 255.255.255.0
 duplex auto
 speed auto
!
interface Serial0/0
 ip address 3.3.3.2 255.255.255.0
```

```
 no fair-queue
!
interface Serial0/1
 ip address 7.7.7.1 255.255.255.0
!
router rip
 redistribute static
 network 7.0.0.0
 network 8.0.0.0
!
router bgp 64500
 no synchronization
 bgp log-neighbor-changes
 network 3.3.3.0 mask 255.255.255.0
 network 7.7.7.0 mask 255.255.255.0
 network 8.8.8.0 mask 255.255.255.0
 network 9.0.0.0
 neighbor 3.3.3.1 remote-as 65500
!
ip classless
ip route 0.0.0.0 0.0.0.0 3.3.3.1
```

（8）路由器 R5 的 BGP 表

```
R5#sh ip bgp
BGP table version is 434, local router ID is 8.8.8.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 1.1.1.0/24       3.3.3.1                  0             0 65500 i
*> 2.2.2.0/24       3.3.3.1                                0 65500 65000 i
*> 3.3.3.0/24       0.0.0.0                  0         32768 i
*> 4.4.4.0/24       3.3.3.1                                0 65500 65000 i
*> 5.5.5.0/24       3.3.3.1                                0 65500 65000 i
*> 6.6.6.0/24       3.3.3.1                                0 65500 65000 i
*> 7.7.7.0/24       0.0.0.0                  0         32768 i
*> 8.8.8.0/24       0.0.0.0                  0         32768 i
*> 9.0.0.0          8.8.8.2                  1         32768 i
```

（9）路由器 R5 的路由表

```
R5#sh ip route
（代码省略）
```

```
Gateway of last resort is 3.3.3.1 to network 0.0.0.0

     1.0.0.0/24 is subnetted, 1 subnets
B       1.1.1.0 [20/0] via 3.3.3.1, 00:52:43
     2.0.0.0/24 is subnetted, 1 subnets
B       2.2.2.0 [20/0] via 3.3.3.1, 00:52:43
     3.0.0.0/24 is subnetted, 1 subnets
C       3.3.3.0 is directly connected, Serial0/0
     4.0.0.0/24 is subnetted, 1 subnets
B       4.4.4.0 [20/0] via 3.3.3.1, 00:52:43
     5.0.0.0/24 is subnetted, 1 subnets
B       5.5.5.0 [20/0] via 3.3.3.1, 00:52:43
     6.0.0.0/24 is subnetted, 1 subnets
B       6.6.6.0 [20/0] via 3.3.3.1, 00:52:43
     7.0.0.0/24 is subnetted, 1 subnets
C       7.7.7.0 is directly connected, Serial0/1
     8.0.0.0/24 is subnetted, 1 subnets
C       8.8.8.0 is directly connected, FastEthernet0/0
R    9.0.0.0/8 [120/1] via 7.7.7.2, 00:00:04, Serial0/1
                       [120/1] via 8.8.8.2, 00:00:00, FastEthernet0/0
S*   0.0.0.0/0 [1/0] via 3.3.3.1
R5#
```

（10）路由器 R7 的配置

```
hostname router7
!
interface Serial1/0
 ip address 4.4.4.1 255.255.255.0
 ip router isis
 clockrate 2000000
!
interface Serial1/2
 ip address 2.2.2.2 255.255.255.0
!
interface Serial1/3
 ip address 5.5.5.1 255.255.255.0
 ip router isis
!
router isis
 default-information originate
 net 49.0002.0000.0000.0007.00
```

```
!
router bgp 65000
 no synchronization
 bgp log-neighbor-changes
 network 2.2.2.0 mask 255.255.255.0
 network 4.4.4.0 mask 255.255.255.0
 network 5.5.5.0 mask 255.255.255.0
 network 6.6.6.0 mask 255.255.255.0
 neighbor 2.2.2.1 rmote-as 65500
```

（11）路由器 R7 的 BGP 表

```
router7#sh ip bgp
BGP table version is 32, local router ID is 5.5.5.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 1.1.1.0/24       2.2.2.1                             0 65500          i
*> 2.2.2.0/24       0.0.0.0                  0         32768            i
*> 3.3.3.0/24       2.2.2.1                             0 65500 64500    i
*> 4.4.4.0/24       0.0.0.0                  0         32768            i
*> 5.5.5.0/24       0.0.0.0                  0         32768            i
*> 6.6.6.0/24       0.0.0.0                 20         32768            i
*> 7.7.7.0/24       2.2.2.1                             0 65500 64500    i
*> 8.8.8.0/24       2.2.2.1                             0 65500 64500    i
*> 9.0.0.0          2.2.2.1                             0 65500 64500    i
router7#
```

（12）路由器 R7 的路由表

```
router7#sh ip rout
（代码省略）

     1.0.0.0/24 is subnetted, 1 subnets
B       1.1.1.0 [20/0] via 2.2.2.1, 01:43:19
     2.0.0.0/24 is subnetted, 1 subnets
C       2.2.2.0 is directly connected, Serial1/2
     3.0.0.0/24 is subnetted, 1 subnets
B       3.3.3.0 [20/0] via 2.2.2.1, 00:40:02
     4.0.0.0/24 is subnetted, 1 subnets
C       4.4.4.0 is directly connected, Serial1/0
     5.0.0.0/24 is subnetted, 1 subnets
```

```
C          5.5.5.0 is directly connected, Serial1/3
       6.0.0.0/24 is subnetted, 1 subnets
i L1       6.6.6.0 [115/20] via 5.5.5.2, Serial1/3
                   [115/20] via 4.4.4.2, Serial1/0
       7.0.0.0/24 is subnetted, 1 subnets
B          7.7.7.0 [20/0] via 2.2.2.1, 00:40:02
       8.0.0.0/24 is subnetted, 1 subnets
B          8.8.8.0 [20/0] via 2.2.2.1, 00:40:02
B      9.0.0.0/8 [20/0] via 2.2.2.1, 00:39:36
router7#
```

（13）路由器 R8 的配置

```
hostname "router8"
!
interface Serial0/0
 ip address 6.6.6.2 255.255.255.0
 ip router isis
!
interface Serial0/2
 ip address 5.5.5.2 255.255.255.0
 ip router isis
 clockrate 72000
!
router isis
 net 49.0002.0000.0000.0008.00
```

（14）路由器 R8 的路由表

```
router8#sh ip rout
（代码省略）

Gateway of last resort is 5.5.5.1 to network 0.0.0.0

       4.0.0.0/24 is subnetted, 1 subnets
i L1       4.4.4.0 [115/20] via 6.6.6.1, Serial0/0
                   [115/20] via 5.5.5.1, Serial0/2
       5.0.0.0/24 is subnetted, 1 subnets
C          5.5.5.0 is directly connected, Serial0/2
       6.0.0.0/24 is subnetted, 1 subnets
C          6.6.6.0 is directly connected, Serial0/0
i*L2 0.0.0.0/0 [115/10] via 5.5.5.1, Serial0/2
router8#
```

（15）路由器 R11 的配置

```
hostname "router11"
!
interface Serial0/0
 ip address 7.7.7.2 255.255.255.0
 clockrate 64000
!
interface FastEthernet0/1
 ip address 9.9.9.1 255.255.255.0
 duplex auto
 speed auto
!
router rip
 network 7.0.0.0
 network 9.0.0.0
```

（16）路由器 R11 的路由表

```
router11#sh ip route
（代码省略）

Gateway of last resort is 7.7.7.1 to network 0.0.0.0

     7.0.0.0/24 is subnetted, 1 subnets
C       7.7.7.0 is directly connected, Serial0/0
R    8.0.0.0/8 [120/1] via 9.9.9.2, 00:00:23, FastEthernet0/1
                [120/1] via 7.7.7.1, 00:00:11, Serial0/0
     9.0.0.0/24 is subnetted, 1 subnets
C       9.9.9.0 is directly connected, FastEthernet0/1
R*   0.0.0.0/0 [120/1] via 7.7.7.1, 00:00:11, Serial0/0
router11#
```

7.13　本 章 小 结

BGP 路由选择协议运行在自治系统之间，为不同自治系统的网络之间通信提供转发数据的路由。因此，BGP 称为外部路由协议。

BGP 协议使用第 4 层的 TCP 协议在对等体之间建立 BGP 会话，端口号为 179，因此，BGP 协议不需要使用其他任何机制来保证它的连接的可靠性。在与对等体建立 BGP 会话时，BGP 使用 6 种邻接状态来描述邻接关系建立的进程。在与邻居同步 BGP 数据库时，BGP 使用 4 种类型不同的消息数据包来控制数据库的同步。

与其他 IGP 路由协议使用度量描述路由的优劣不同，BGP 协议使用一系列路径属性描述

路由的优劣。BGP 路由器通过比较这些路径属性值选择最佳路由，然后把最佳路由安装到路由表中。在 Cisco 路由器上，选择最佳路由的顺序如 7.8 节所述。

BGP 协议运行遵循两个规则，一个是同步规则，另一个是 Split Horizon 规则。

同步规则规定，在 BGP 同步功能开启时，BGP 路由器不会使用或向它的 EBGP 邻居通告一条来自 IBGP 对等体的路由，除非这条路由也能通过 IGP 学习到，即 BGP 库和 IGP 库是同步的。可以在路由器上关闭同步功能以绕开这个规则的限制，但是，为了避免出现路由黑洞，只有在 BGP 转接路径上的所有路由器都运行了 BGP 时，关闭同步功能才是安全的。

Split Horizon 规则规定，BGP 路由器不会把来自某个 IBGP 对等体的路由信息再转发给其他 IBGP 对等体。因此，为了使 BGP 路由信息能够在各个路由器之间扩散，IBGP 对等体之间就必须建立全互连的 IBGP 会话。随着路由器数量的增加，TCP 会话数量也将急速增加，并因此带来 TCP 流量的增加。使用路由反射器技术既可以避免建立全互连的 IBGP 会话，又能保证路由信息的顺利扩散。

在路由器上配置 BGP 协议的基本工作有两个，一个是在路由器上启动 BGP 协议并配置对等体；另一个是向 BGP 进程中通告路由。启动 BGP 的命令是“router bgp *AS-number*”；配置对等体的命令是“neighbor…remote-as…”。可以使用“network”把直连网络或 IGP 路由通告到 BGP 进程，也可以使用“aggregate-address”命令向 BGP 进程通告一条聚合的 BGP 路由。

另外，本章还介绍了如下 BGP 技术：

- 使用环回接口建立 BGP 会话；
- 更改下一跳属性值；
- 配置对等体组；
- 应用团体属性；
- 更改权重属性值；
- 更改本地优先属性值；
- 其他一些管理 BGP 运行的命令。

最后，本章介绍了 BGP 多宿主连接。

第 8 章　多协议 BGP 之 IPv6 BGP

多协议 BGP（Multiprotocol BGP）是增强型的 BGP 路由协议，这种增强使得 BGP 协议不仅可以携带 IPv4 地址族信息，还可以携带其他网络层协议地址族（Address Family）信息，例如，IPv6 地址族信息、IPv6 组播路由等。

本章只讨论 IPv6 单播特性的 MBGP。

8.1　BGP-4 的多协议扩展

当前的 BGP-4 仅具备携带 IPv4 路由信息的能力，对 BGP-4 的多协议扩展（RFC2858）使得它可以携带多种网络层协议信息，比如 IPv6 等。这种扩展特性是向后兼容的，不识别扩展属性的路由器和识别扩展属性的路由器可以共存在一个网络中，不识别扩展属性的路由器忽略扩展属性即可。

在 BGP-4 中，携带的与 IPv4 特别相关的信息包括：

① 以 IPv4 地址表达的下一跳（Next_hop）属性；

② 包含 IPv4 地址的聚合者（Aggregator）属性；

③ 以 IPv4 地址前缀表达的网络层可达信息（NLRI）。

所以，为了让 BGP-4 支持多种网络层协议，需要在以下两方面扩展 BGP-4 的功能：

① 特定网络层协议与下一跳相关联的能力；

② 特定网络层协议与 NLRI 相关联的能力。

为了向后兼容，同时简化 BGP-4 对多协议的支持，引入了两个新属性：

① 多协议可达 NLRI（MP_REACH_NLRI），用来携带可达的目标网络、下一跳地址信息和路由器的 SNPA 地址。

② 多协议不可达 NLRI（MP_UNREACH_NLRI），用来携带不可达的（撤销的）目标网络信息。

这两个属性都是可选不传递属性，不会传递给 BGP 对等体。这样，不识别多协议属性的路由器只需忽略它们即可。

8.1.1　多协议可达 NLRI

多协议可达 NLRI 类型代码为 14，结构如图 8-1 所示。

- Address Family Identifier：地址族标识，2 字节。
- Subsequent Address Family Identifier：地址族附加信息，比如携带单播信息、组播信息或二者信息都携带，1 字节。
- Length of Next Hop Network Address：下一跳地址的长度，1 字节。

- Network Address of Next Hop：下一跳地址，变长。
- Number of SNPA：SNPA 地址数量，1 字节。
- Length of first SNPA：第一个 SNPA 地址长度，1 字节。
- First SNPA：第一个 SNPA 地址。
- Network Layer Reachability Information：网络层可达信息。

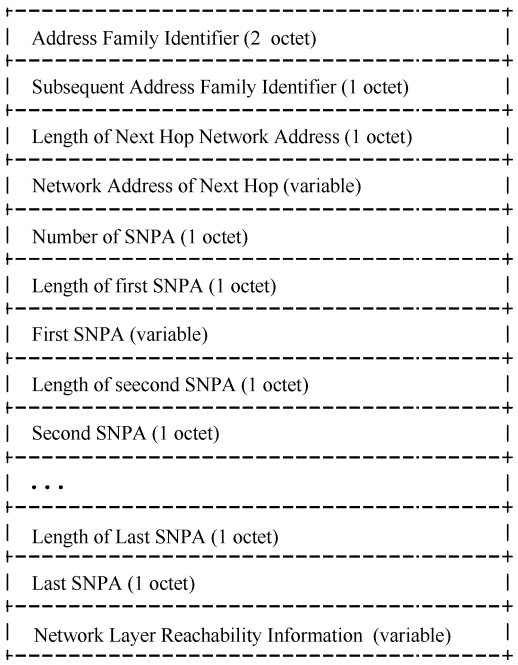
Address Family Identifier (2 octet)
Subsequent Address Family Identifier (1 octet)
Length of Next Hop Network Address (1 octet)
Network Address of Next Hop (variable)
Number of SNPA (1 octet)
Length of first SNPA (1 octet)
First SNPA (variable)
Length of seecond SNPA (1 octet)
Second SNPA (1 octet)
. . .
Length of Last SNPA (1 octet)
Last SNPA (1 octet)
Network Layer Reachability Information (variable)

图 8-1　多协议可达 NLRI 报文结构

8.1.2　多协议不可达 NLRI

多协议不可达 NLRI 类型代码为 15，结构如图 8-2 所示。

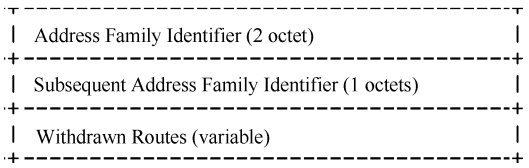

图 8-2　多协议不可达 NLRI 报文结构

- Withdrawn Routes：撤销的路由，变长。

8.1.3　NLRI 的结构

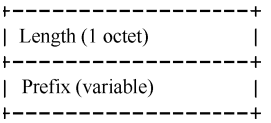

图 8-3　NLRI 结构

NLRI 使用长度 / 前缀组合表示。重复使用这样的组合，可以携带多个路由前缀，结构如图 8-3 所示。

前缀（Prefix）部分必须是整字节，不够整字节的需要填充至整字节长。

8.1.4　多协议特性通告

如果路由器支持多协议，它在 OPEN 报文的可选参数字段通告其能力（Capability）特性。如果对方路由器不支持多协议，则使用 Notification 回应。BGP-4 能力通告（Capabilities Advertisement with BGP-4）由 RFC2842 定义。

多协议能力通告使用［类型，长度，值］的结构，在可选参数字段携带。

8.2　配置支持 IPv6 单播的 MBGP

基本配置命令和 IPv4 下配置 BGP 的命令类似。本例启用 BGP 仅通告 IPv6 路由，因此需要关闭 IPv4 单播路由的通告功能。默认地，IPv4 单播路由通告功能是开启的。关闭命令是 no bgp default ipv4-unicast，和 IPv4 下配置 BGP 不同的是，为支持 IPv6 地址族信息，需要在 IPv6 地址族模式下激活邻居。

8.2.1　基本配置命令

下面以图 8-4 的拓扑为例，介绍 MBGP 的基本配置命令。

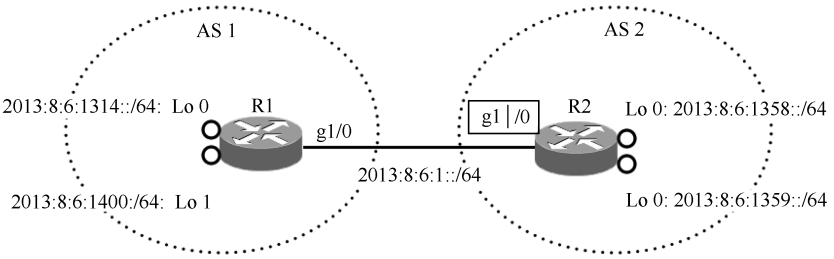

图 8-4　介绍 MBGP 配置的拓扑

先查看一下 2 个 g1/0 接口的 IPv6 地址。

```
R1#sh ipv6 int g1/0
GigabitEthernet1/0 is up, line protocol is up
  IPv6 is enabled, link-local address is FE80::C800:8FF:FE5B:1C
  No Virtual link-local address(es):
  Global unicast address(es):
    2013:8:6:1:C800:8FF:FE5B:1C, subnet is 2013:8:6:1::/64 [EUI]
……
```

```
R2#sh ipv6 int g1/0
GigabitEthernet1/0 is up, line protocol is up
  IPv6 is enabled, link-local address is FE80::C801:8FF:FE5B:1C
  No Virtual link-local address(es):
```

```
    Global unicast address(es):
      2013:8:6:1:C801:8FF:FE5B:1C, subnet is 2013:8:6:1::/64 [EUI]
……
```

在 R1 上，把两个 Loopback 接口代表的网段通过 MBGP 通告给 R2，同时启用了 graceful restart 特性。

（1）R1 的相关配置

```
hostname R1
!
ip cef
ipv6 unicast-routing
ipv6 cef
!
interface Loopback0
 no ip address
 ipv6 address 2013:8:6:1314::/64 eui-64
 ipv6 enable
!
interface Loopback1
 no ip address
 ipv6 address 2013:8:6:1400::/64 eui-64
 ipv6 enable
!
interface GigabitEthernet1/0
 no ip address
 negotiation auto
 ipv6 address 2013:8:6:1::/64 eui-64
 ipv6 enable
!
router bgp 1
 bgp router-id 1.1.1.1
 bgp log-neighbor-changes
 bgp graceful-restart restart-time 120
 bgp graceful-restart stalepath-time 360
 bgp graceful-restart extended
 no bgp default ipv4-unicast
 neighbor 2013:8:6:1:C801:8FF:FE5B:1C remote-as 2
 !
 address-family ipv4
 exit-address-family
 !
```

```
 address-family ipv6
  network 2013:8:6:1314::/64
  network 2013:8:6:1400::/64
  neighbor 2013:8:6:1:C801:8FF:FE5B:1C activate
 exit-address-family
```

在 R2 上，把两个 Loopback 接口代表的网段通过 MBGP 通告给 R1，同时启用了 graceful restart 特性。

（2）R2 的相关配置

```
hostname R2
!
ip cef
ipv6 unicast-routing
ipv6 cef
!
interface Loopback0
 no ip address
 ipv6 address 2013:8:6:1358::/64 eui-64
 ipv6 enable
!
interface Loopback1
 no ip address
 ipv6 address 2013:8:6:1359::/64 eui-64
 ipv6 enable
!
interface GigabitEthernet1/0
 no ip address
 negotiation auto
 ipv6 address 2013:8:6:1::/64 eui-64
 ipv6 enable
!
router bgp 2
 bgp router-id 2.2.2.2
 bgp log-neighbor-changes
 bgp graceful-restart restart-time 120
 bgp graceful-restart stalepath-time 360
 bgp graceful-restart extended
 no bgp default ipv4-unicast
 neighbor 2013:8:6:1:C800:8FF:FE5B:1C remote-as 1
 !
 address-family ipv4
 exit-address-family
```

```
!
address-family ipv6
  network 2013:8:6:1358::/64
  network 2013:8:6:1359::/64
  neighbor 2013:8:6:1:C800:8FF:FE5B:1C activate
exit-address-family
```

8.2.2　检查运行状态

（1）查看邻居关系

配置完成后，查看一下邻居关系是否建立起来了。

```
R1#show bgp ipv6 unicast neighbors
BGP neighbor is 2013:8:6:1:C801:8FF:FE5B:1C,    remote AS 2, external link
  BGP version 4, remote router ID 2.2.2.2
  BGP state = Established, up for 00:31:54
……..
Local host: 2013:8:6:1:C800:8FF:FE5B:1C, Local port: 29292
Foreign host: 2013:8:6:1:C801:8FF:FE5B:1C, Foreign port: 179
```

Established 表示 TCP 会话已经建立起来了。

（2）查看 IPv6 BGP 库

```
R2#sh bgp ipv6 unicast
BGP table version is 9, local router ID is 2.2.2.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
              x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

     Network          Next Hop            Metric  LocPrf  Weight  Path
 *>  2013:8:6:1314::/64
                      2013:8:6:1:C800:8FF:FE5B:1C
                                           0               0       1     i
 *>  2013:8:6:1358::/64
                      ::                   0               32768         i
 *>  2013:8:6:1359::/64
                      ::                   0               32768         i
 *>  2013:8:6:1400::/64
                      2013:8:6:1:C800:8FF:FE5B:1C
                                           0               0       1     i
R2#
```

（3）查看路由表

```
R1#sh ipv6 route bgp
IPv6 Routing Table - default - 9 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
       B - BGP, HA - Home Agent, MR - Mobile Router, R - RIP
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       D - EIGRP, EX - EIGRP external, ND - ND Default, NDp - ND Prefix
       DCE - Destination, NDr - Redirect
B     2013:8:6:1358::/64 [20/0]
       via FE80::C801:8FF:FE5B:1C, GigabitEthernet1/0
B     2013:8:6:1359::/64 [20/0]
       via FE80::C801:8FF:FE5B:1C, GigabitEthernet1/0
R1#
```

8.3　把 IPv6 路由重分布到 mBGP

如同 IPv4 环境一样，需要把 IPv6 的 IGP 路由信息，通过路由重分布技术注入到 mBGP 进程，从而传递到其他 AS。图 8-5 的拓扑是用来演示 IPv6 路由重分布的。

在 AS 100 中的路由器 R1 和 R3 运行 RIPng；在 AS 200 中的路由器 R2 和 R4 运行 RIPng。

路由器 R1 和 R2 运行 MBGP，并把 RIPng 的路由重分布到 IPv6 地址族中。在 R1 和 R2 路由器上分别向自己的 AS 内产生默认路由。

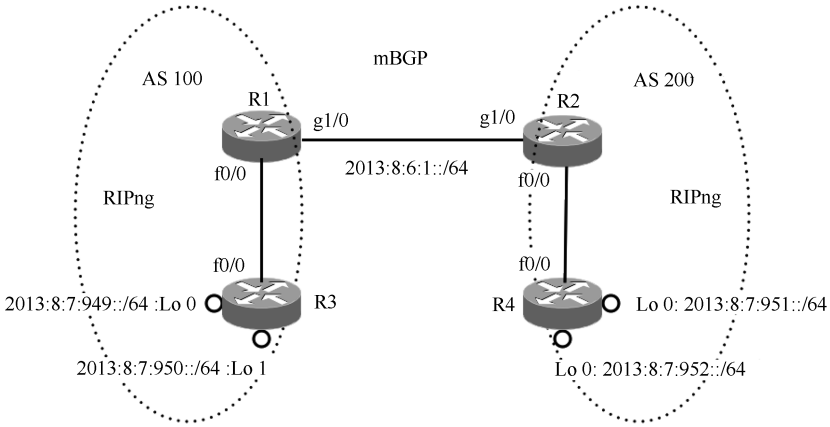

图 8-5　IPv6 路由重分布

（1）路由器 R1 的相关配置

```
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
```

```
 ipv6 enable
 ipv6 rip pro1 enable
 ipv6 rip pro1 default-information originate metric 2
!
!
interface GigabitEthernet1/0
 no ip address
 negotiation auto
 ipv6 address 2013:8:6:1::/64 eui-64
 ipv6 enable
!
router bgp 100
 bgp router-id 1.1.1.1
 bgp log-neighbor-changes
 bgp graceful-restart restart-time 120
 bgp graceful-restart stalepath-time 360
 bgp graceful-restart
 no bgp default ipv4-unicast
 neighbor 2013:8:6:1:C801:AFF:FEBA:1C remote-as 200
 !
 address-family ipv4
 exit-address-family
 !
 address-family ipv6
  redistribute rip pro1
  neighbor 2013:8:6:1:C801:AFF:FEBA:1C activate
 exit-address-family
!
!
ipv6 router rip pro1
!
```

（2）路由器 R2 的相关配置

```
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 enable
 ipv6 rip pro2 enable
 ipv6 rip pro2 default-information originate metric 2
!
```

```
!
interface GigabitEthernet1/0
 no ip address
 negotiation auto
 ipv6 address 2013:8:6:1::/64 eui-64
 ipv6 enable
!
router bgp 200
 bgp router-id 2.2.2.2
 bgp log-neighbor-changes
 bgp graceful-restart restart-time 120
 bgp graceful-restart stalepath-time 360
 bgp graceful-restart
 no bgp default ipv4-unicast
 neighbor 2013:8:6:1:C800:AFF:FEBA:1C remote-as 100
 !
 address-family ipv4
 exit-address-family
 !
 address-family ipv6
  redistribute rip pro2
  neighbor 2013:8:6:1:C800:AFF:FEBA:1C activate
 exit-address-family
!
ipv6 router rip pro2
```

（3）路由器 R3 的相关配置

```
interface Loopback0
 no ip address
 ipv6 address 2013:8:7:949::/64 eui-64
 ipv6 enable
 ipv6 rip pro3 enable
!
interface Loopback1
 no ip address
 ipv6 address 2013:8:7:950::/64 eui-64
 ipv6 enable
 ipv6 rip pro3 enable
!
interface FastEthernet0/0
 no ip address
```

```
 duplex auto
 speed auto
 ipv6 enable
 ipv6 rip pro3 enable
!
ipv6 router rip pro3
```

（4）路由器 R4 的相关配置

```
interface Loopback0
 no ip address
 ipv6 address 2013:8:7:951::/64 eui-64
 ipv6 enable
 ipv6 rip pro4 enable
!
interface Loopback1
 no ip address
 ipv6 address 2013:8:7:952::/64 eui-64
 ipv6 enable
 ipv6 rip pro4 enable
!
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 enable
 ipv6 rip pro4 enable
!
!
ipv6 router rip pro4
```

（5）路由器 R1 的路由表

```
R1#sh ipv6 route
IPv6 Routing Table - default - 7 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
       B - BGP, HA - Home Agent, MR - Mobile Router, R - RIP
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       D - EIGRP, EX - EIGRP external, ND - ND Default, NDp - ND Prefix
       DCE - Destination, NDr - Redirect
C      2013:8:6:1::/64 [0/0]
        via GigabitEthernet1/0, directly connected
```

```
L     2013:8:6:1:C800:AFF:FEBA:1C/128 [0/0]
        via GigabitEthernet1/0, receive
R     2013:8:7:949::/64 [120/2]
        via FE80::C803:AFF:FEBC:8, FastEthernet0/0
R     2013:8:7:950::/64 [120/2]
        via FE80::C803:AFF:FEBC:8, FastEthernet0/0
B     2013:8:7:951::/64 [20/2]
        via FE80::C801:AFF:FEBA:1C, GigabitEthernet1/0
B     2013:8:7:952::/64 [20/2]
        via FE80::C801:AFF:FEBA:1C, GigabitEthernet1/0
L     FF00::/8 [0/0]
        via Null0, receive
R1#
```

（6）路由器 R3 的路由表

```
R3#sh ipv6 route
IPv6 Routing Table - default - 6 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
           B - BGP, HA - Home Agent, MR - Mobile Router, R - RIP
           I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
           D - EIGRP, EX - EIGRP external, ND - ND Default, NDp - ND Prefix
           DCE - Destination, NDr - Redirect
R     ::/0 [120/3]
        via FE80::C800:AFF:FEBA:8, FastEthernet0/0
C     2013:8:7:949::/64 [0/0]
        via Loopback0, directly connected
L     2013:8:7:949:C803:AFF:FEBC:8/128 [0/0]
        via Loopback0, receive
C     2013:8:7:950::/64 [0/0]
        via Loopback1, directly connected
L     2013:8:7:950:C803:AFF:FEBC:8/128 [0/0]
        via Loopback1, receive
L     FF00::/8 [0/0]
        via Null0, receive
R3#
```

（7）路由器 R2 的路由表

```
R2#sh ipv6 route
IPv6 Routing Table - default - 7 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
```

```
            B - BGP, HA - Home Agent, MR - Mobile Router, R - RIP
            I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
            D - EIGRP, EX - EIGRP external, ND - ND Default, NDp - ND Prefix
            DCE - Destination, NDr - Redirect
C     2013:8:6:1::/64 [0/0]
       via GigabitEthernet1/0, directly connected
L     2013:8:6:1:C801:AFF:FEBA:1C/128 [0/0]
       via GigabitEthernet1/0, receive
B     2013:8:7:949::/64 [20/2]
       via FE80::C800:AFF:FEBA:1C, GigabitEthernet1/0
B     2013:8:7:950::/64 [20/2]
       via FE80::C800:AFF:FEBA:1C, GigabitEthernet1/0
R     2013:8:7:951::/64 [120/2]
       via FE80::C804:AFF:FEBC:8, FastEthernet0/0
R     2013:8:7:952::/64 [120/2]
       via FE80::C804:AFF:FEBC:8, FastEthernet0/0
L     FF00::/8 [0/0]
       via Null0, receive
R2#
```

（8）路由器 R4 的路由表

```
R4#sh ipv6 route rip
IPv6 Routing Table - default - 6 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
            B - BGP, HA - Home Agent, MR - Mobile Router, R - RIP
            I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
            D - EIGRP, EX - EIGRP external, ND - ND Default, NDp - ND Prefix
            DCE - Destination, NDr - Redirect
R     ::/0 [120/3]
       via FE80::C801:AFF:FEBA:8, FastEthernet0/0
R4#
```

8.4　IPv6 邻居交换 IPv4 路由

图 8-6 模拟的是 2 个 IPv4 的路由域被一个 IPv6 路由域隔离，R1 和 R2 只配置了 IPv6 地址，并建立了 iBGP 邻居关系。在这种情况下，也可以让它们通过 MBGP 交换 IPv4 的路由信息。

实现该目标只需在 IPv4 地址族下，用 IPv6 地址激活邻居关系，并把 RIP 路由信息重分布到 IPv4 地址族，配置如下所述。

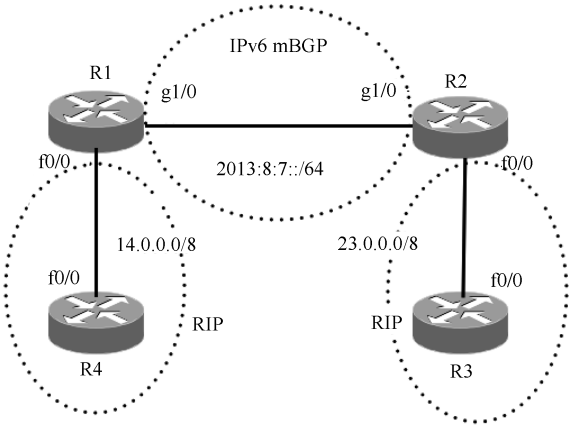

图 8-6　跨 IPv6 网络交换 IPv4 路由

（1）路由器 R1 的配置

```
hostname R1
!
ip cef
ipv6 unicast-routing
ipv6 cef
!
!
interface Loopback0
 ip address 10.0.0.1 255.255.255.255
!
interface FastEthernet0/0
 ip address 14.0.0.1 255.0.0.0
 duplex auto
 speed auto
!
!
interface GigabitEthernet1/0
 no ip address
 negotiation auto
 ipv6 address 2013:8:7::1/64
 ipv6 enable
!
router rip
 version 2
 network 14.0.0.0
 default-information originate
 no auto-summary
```

```
!
router bgp 10
 bgp router-id 1.1.1.1
 bgp log-neighbor-changes
 neighbor 2013:8:7::2 remote-as 10
 !
 address-family ipv4
  redistribute rip
  neighbor 2013:8:7::2 activate
 exit-address-family
!
```

（2）路由器 R2 的配置

```
hostname R2
!
!
ip cef
ipv6 unicast-routing
ipv6 cef

!
interface Loopback0
 ip address 20.0.0.1 255.255.255.255
!
interface FastEthernet0/0
 ip address 23.0.0.2 255.0.0.0
 duplex auto
 speed auto
!
interface GigabitEthernet1/0
 no ip address
 negotiation auto
 ipv6 address 2013:8:7::2/64
 ipv6 enable
!
router rip
 version 2
 network 23.0.0.0
 default-information originate
 no auto-summary
!
```

```
router bgp 10
 bgp router-id 2.2.2.2
 bgp log-neighbor-changes
 neighbor 2013:8:7::1 remote-as 10
 !
 address-family ipv4
  redistribute rip
  neighbor 2013:8:7::1 activate
 exit-address-family
!
!
!
route-map change-next-hop permit 10
 set ip next-hop 20.0.0.1
!
```

（3）查看 BGP 数据库

配置完成后，查看一下 BGP 数据库：

```
R1#sh ip bgp
BGP table version is 4, local router ID is 1.1.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
              x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

     Network          Next Hop            Metric LocPrf Weight Path
 *>  14.0.0.0         0.0.0.0                  0         327 68 ?
 * i 23.0.0.0         32.19.0.8                0    100  0    ?
R1#
```

```
R2#sh ip bgp
BGP table version is 4, local router ID is 2.2.2.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
              x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

     Network          Next Hop            Metric LocPrf Weight Path
 * i 14.0.0.0         32.19.0.8                0    100  0    ?
```

```
 *>  23.0.0.0          0.0.0.0                  0         32768      ?
R2#
```

数据库显示，须由 iBGP 传送来的路由都不可用（没有大于号标示），原因是下一跳地址（32.19.0.8）在 R1 和 R2 上都是不可达的。

为了使经由 IPv6 BGP 邻居传送来的 IPv4 路由可用，只需要保证下一跳可达即可，比如可以使用 route-map 修改下一跳属性值。

在 R1 和 R2 上分别配置一个名字叫 change-next-hop 的 route-map，把下一跳地址更改为它们的 loopback 接口地址。

注意：这里使用 Loopback 接口的地址只是为了演示如何保证下一跳可达的方法，并无特殊作用，也不能使得 2 个 IPv4 路由域通信。

（4）修改路由器 R1 的配置

```
router bgp 10
 bgp router-id 1.1.1.1
 bgp log-neighbor-changes
 neighbor 2013:8:7::2 remote-as 10
 !
 address-family ipv4
  redistribute rip
  neighbor 2013:8:7::2 activate
  neighbor 2013:8:7::2 route-map change-next-hop in
 exit-address-family
!
route-map change-next-hop permit 10
 set ip next-hop 10.0.0.1
```

（5）修改路由器 R2 的配置

```
router bgp 10
 bgp router-id 2.2.2.2
 bgp log-neighbor-changes
 neighbor 2013:8:7::1 remote-as 10
 !
 address-family ipv4
  redistribute rip
  neighbor 2013:8:7::1 activate
  neighbor 2013:8:7::1 route-map change-next-hop in
 exit-address-family
!
route-map change-next-hop permit 10
 set ip next-hop 20.0.0.1
```

提示：修改配置后，需要重置 BGP 进程，使新的配置命令生效。

（6）再次查看 2 台路由器的 BGP 数据库和路由表

① R1 的状态：

```
R1#sh ip bgp
BGP table version is 5, local router ID is 1.1.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
              x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

     Network          Next Hop            Metric LocPrf Weight Path
 *>  14.0.0.0         0.0.0.0                  0         32768 ?
 *>i 23.0.0.0         10.0.0.1                 0    100  0     ?
R1#
R1#
R1#sh ip route
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      10.0.0.0/32 is subnetted, 1 subnets
C        10.0.0.1 is directly connected, Loopback0
      14.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
C        14.0.0.0/8 is directly connected, FastEthernet0/0
L        14.0.0.1/32 is directly connected, FastEthernet0/0
B     23.0.0.0/8 [200/0] via 10.0.0.1, 00:00:28
R1#
```

② R2 的状态：

```
R2#sh ip bgp
BGP table version is 5, local router ID is 2.2.2.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
```

```
                x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

     Network          Next Hop            Metric  LocPrf  Weight  Path
 *>i 14.0.0.0         20.0.0.1                 0     100       0  ?
 *>  23.0.0.0         0.0.0.0                  0               32768  ?
R2#
R2#
R2#sh ip route
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

B       14.0.0.0/8 [200/0] via 20.0.0.1, 00:00:15
       20.0.0.0/32 is subnetted, 1 subnets
C          20.0.0.1 is directly connected, Loopback0
       23.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
C          23.0.0.0/8 is directly connected, FastEthernet0/0
L          23.0.0.2/32 is directly connected, FastEthernet0/0
R2#
```

8.5　IPv4 邻居交换 IPv6 路由

图 8-7 模拟的是 2 个 IPv6 的路由域被一个 IPv4 路由域隔离。R1 和 R2 只配置了 IPv4 地址，并建立了 iBGP 邻居关系。在这种情况下，也可以让它们通过 MBGP 交换 IPv6 的路由信息。

（1）路由器 R1 的相关配置

```
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 enable
 ipv6 rip pro1 enable
```

```
  ipv6 rip pro1 default-information originate
!
interface GigabitEthernet1/0
  ip address 12.0.0.1 255.0.0.0
  negotiation auto
!
router bgp 10
  bgp router-id 1.1.1.1
  bgp log-neighbor-changes
  no bgp default ipv4-unicast
  neighbor 12.0.0.2 remote-as 10
  !
  address-family ipv4
    neighbor 12.0.0.2 activate
  exit-address-family
  !
  address-family ipv6
    redistribute rip pro1
    neighbor 12.0.0.2 activate
  exit-address-family
!
!
ipv6 router rip pro1
```

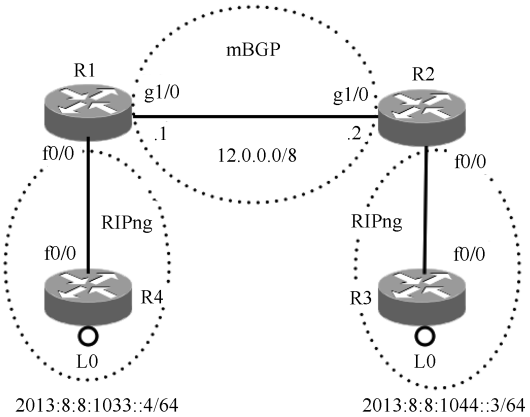

图 8-7　跨 IPv4 网络交换 IPv6 路由

（2）路由器 R2 的相关配置

```
interface FastEthernet0/0
  no ip address
  duplex auto
  speed auto
```

```
 ipv6 enable
 ipv6 rip pro1 enable
 ipv6 rip pro1 default-information originate
!
interface GigabitEthernet1/0
 ip address 12.0.0.2 255.0.0.0
 negotiation auto
!
router bgp 10
 bgp router-id 2.2.2.2
 bgp log-neighbor-changes
 no bgp default ipv4-unicast
 neighbor 12.0.0.1 remote-as 10
 !
 address-family ipv4
  neighbor 12.0.0.1 activate
 exit-address-family
 !
 address-family ipv6
  redistribute rip pro1
  neighbor 12.0.0.1 activate
 exit-address-family
!
!
ipv6 router rip pro1
```

（3）在 R2 上查看 IPv6 BGP 数据库

```
R2#sh bgp ipv6 unicast
BGP table version is 2, local router ID is 2.2.2.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
              x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

     Network          Next Hop            Metric  LocPrf  Weight  Path
 * i 2013:8:8:1033::/64
                      ::FFFF:12.0.0.1          2     100       0  ?
R2#
```

由于下一跳不可达，所以该路由不是最佳（> best）路由，不能写入路由表。

为保证下一跳可达，可以使用 route-map 修改下一跳路由。

8.6　示例：调整本地优先值

本示例的拓扑如图 8-8 所示。

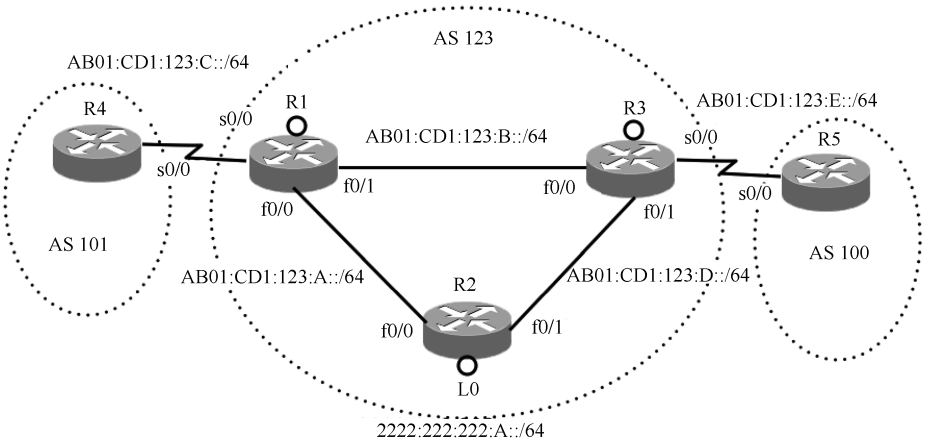

图 8-8

在路由器 R4 和 R5 上都向 BGP 注入了相同的 3 条路由：

BC01:BC1:10:A::/64

BC02:BC1:11:A::/64

BC03:BC1:12:A::/64

在路由器 R3 上把路由 BC01:BC1:10:A::/64 的本地优先值修改为 500（默认值=100），其他 2 条路由的本地优先值不更改。在 R1 上，这 3 条路由的本地优先值都是默认值。

（1）路由器 R1 的配置清单

```
interface Loopback0
 no ip address
 ipv6 address 1111:111:111:A::/64 eui-64
 ipv6 enable
 ipv6 ospf 10 area 0
!
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 address AB01:CD1:123:A::/64 eui-64
 ipv6 enable
 ipv6 ospf 10 area 0
!
interface Serial0/0
```

```
 no ip address
 ipv6 address AB01:CD1:123:C::/64 eui-64
 ipv6 enable
 clock rate 2000000
!
interface FastEthernet0/1
 no ip address
 duplex auto
 speed auto
 ipv6 address AB01:CD1:123:B::/64 eui-64
 ipv6 enable
 ipv6 ospf 10 area 0
!
ipv6 router ospf 10
 router-id 1.1.1.1
 log-adjacency-changes
 redistribute connected route-map CONNECTED
!
route-map CONNECTED permit 10
 match interface Serial0/0
!
router bgp 123
 no synchronization
 bgp router-id 1.1.1.1
 no bgp default ipv4-unicast
 bgp log-neighbor-changes
 neighbor 2222:222:222:A:C602:3FF:FEF0:0 remote-as 123
 neighbor 2222:222:222:A:C602:3FF:FEF0:0 update-source Loopback0
 neighbor 3333:333:333:A:C603:3FF:FEF0:0 remote-as 123
 neighbor 3333:333:333:A:C603:3FF:FEF0:0 update-source Loopback0
 neighbor AB01:CD1:123:C:C604:16FF:FE98:0 remote-as 101
!
 address-family ipv6
  neighbor 2222:222:222:A:C602:3FF:FEF0:0 activate
  neighbor 2222:222:222:A:C602:3FF:FEF0:0 next-hop-self
  neighbor 3333:333:333:A:C603:3FF:FEF0:0 activate
  neighbor 3333:333:333:A:C603:3FF:FEF0:0 next-hop-self
  neighbor AB01:CD1:123:C:C604:16FF:FE98:0 activate
exit-address-family
```

（2）路由器 R2 的配置清单

```
interface Loopback0
 no ip address
 ipv6 address 2222:222:222:A::/64 eui-64
 ipv6 enable
 ipv6 ospf 10 area 0
!
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 address AB01:CD1:123:A::/64 eui-64
 ipv6 ospf 10 area 0
!
interface FastEthernet0/1
 no ip address
 duplex auto
 speed auto
 ipv6 address AB01:CD1:123:D::/64 eui-64
 ipv6 enable
 ipv6 ospf 10 area 0
!
ipv6 router ospf 10
 router-id 2.2.2.2
 log-adjacency-changes
!
router bgp 123
 no synchronization
 bgp router-id 2.2.2.2
 no bgp default ipv4-unicast
 bgp log-neighbor-changes
 neighbor 1111:111:111:A:C601:3FF:FEF0:0 remote-as 123
 neighbor 1111:111:111:A:C601:3FF:FEF0:0 update-source Loopback0
 neighbor 3333:333:333:A:C603:3FF:FEF0:0 remote-as 123
 neighbor 3333:333:333:A:C603:3FF:FEF0:0 update-source Loopback0
 no auto-summary
 !
 address-family ipv6
  neighbor 1111:111:111:A:C601:3FF:FEF0:0 activate
  neighbor 3333:333:333:A:C603:3FF:FEF0:0 activate
 exit-address-family
```

（3）路由器 R3 的配置清单

```
interface Loopback0
 no ip address
 ipv6 address 3333:333:333:A::/64 eui−64
 ipv6 enable
 ipv6 ospf 10 area 0
!
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 address AB01:CD1:123:B::/64 eui−64
 ipv6 enable
 ipv6 ospf 10 area 0
!
interface Serial0/0
 no ip address
 ipv6 address AB01:CD1:123:E::/64 eui−64
 ipv6 enable
 clock rate 2000000
!
interface FastEthernet0/1
no ip address
 duplex auto
 speed auto
 ipv6 address AB01:CD1:123:D::/64 eui−64
 ipv6 ospf 10 area 0
!
ipv6 router ospf 10
 router−id 3.3.3.3
 log−adjacency−changes
 redistribute connected route−map CONNECTED
!
router bgp 123
 no synchronization
 no bgp default ipv4−unicast
 bgp router−id 3.3.3.3
 bgp log−neighbor−changes
 neighbor 1111:111:111:A:C601:3FF:FEF0:0 remote−as 123
 neighbor 1111:111:111:A:C601:3FF:FEF0:0 update−source Loopback0
 neighbor 2222:222:222:A:C602:3FF:FEF0:0 remote−as 123
```

```
  neighbor 2222:222:222:A:C602:3FF:FEF0:0 update-source Loopback0
  neighbor AB01:CD1:123:E:C605:16FF:FE98:0 remote-as 100
  no auto-summary
  !
  address-family ipv6
   neighbor 1111:111:111:A:C601:3FF:FEF0:0 activate
   neighbor 1111:111:111:A:C601:3FF:FEF0:0 next-hop-self
   neighbor 1111:111:111:A:C601:3FF:FEF0:0 route-map LOCAL_PREF out
   neighbor 2222:222:222:A:C602:3FF:FEF0:0 activate
   neighbor 2222:222:222:A:C602:3FF:FEF0:0 next-hop-self
   neighbor 2222:222:222:A:C602:3FF:FEF0:0 route-map LOCAL_PREF out
   neighbor AB01:CD1:123:E:C605:16FF:FE98:0 activate
  exit-address-family
!
ipv6 prefix-list 10 seq 5 permit BC01:BC1:10:A::/64
!
route-map LOCAL_PREF permit 10
 match ipv6 address prefix-list 10
 set local-preference 500
!
route-map LOCAL_PREF permit 20
!
route-map CONNECTED permit 10
 match interface Serial0/0
```

（4）路由器 R4 的配置清单

```
interface Loopback10
 no ip address
 ipv6 address BC01:BC1:10:A::/64 eui-64
 ipv6 enable
!
interface Loopback11
 no ip address
 ipv6 address BC02:BC1:11:A::/64 eui-64
 ipv6 enable
!
interface Loopback12
no ip address
ipv6 address BC03:BC1:12:A::/64 eui-64
 ipv6 enable
!
```

```
interface Serial0/0
 no ip address
 ipv6 address AB01:CD1:123:C::/64 eui-64
 ipv6 enable
 clock rate 2000000
!
router bgp 101
 bgp router-id 4.4.4.4
 no bgp default ipv4-unicast
 bgp log-neighbor-changes
 neighbor AB01:CD1:123:C:C601:3FF:FEF0:0 remote-as 123
 !
 address-family ipv6
  neighbor AB01:CD1:123:C:C601:3FF:FEF0:0 activate
  network BC01:BC1:10:A::/64
  network BC02:BC1:11:A::/64
  network BC03:BC1:12:A::/64
 exit-address-family
```

（5）路由器 R5 的配置清单

```
interface Loopback10
 no ip address
 ipv6 address BC01:BC1:10:A::/64 eui-64
 ipv6 enable
!
interface Loopback11
 no ip address
 ipv6 address BC02:BC1:11:A::/64 eui-64
 ipv6 enable
!
interface Loopback12
 no ip address
 ipv6 address BC03:BC1:12:A::/64 eui-64
 ipv6 enable
!
interface Serial0/0
 no ip address
 ipv6 address AB01:CD1:123:E::/64 eui-64
 ipv6 enable
 clock rate 2000000
!
```

```
router bgp 100
 bgp router-id 5.5.5.5
 no bgp default ipv4-unicast
 bgp log-neighbor-changes
 neighbor AB01:CD1:123:E:C603:3FF:FEF0:0 remote-as 123
 !
 address-family ipv6
  neighbor AB01:CD1:123:E:C603:3FF:FEF0:0 activate
  network BC01:BC1:10:A::/64
  network BC02:BC1:11:A::/64
  network BC03:BC1:12:A::/64
 exit-address-family
```

8.7　本 章 小 结

为了使目前的 BGP（版本 4 BGP）支持支持更多的路由信息，比如 IPv6、VPNv4、组播等路由信息，必须对其进行扩展。扩展的目的是增加支持多协议的可达与不可达 NLRI 属性，使其携带多协议路由信息以及将要撤销的多协议路由信息。

本章介绍了 NLRI 扩展及报文结构。本章主要介绍的是支持 IPv6 单播路由的多协议 BGP 特性，并举例介绍了配置支持 IPv6 单播 MBGP 的基本命令。本章还特别介绍了如何实现 IPv4 BGP 邻居传递 IPv6 路由信息和 IPv6 BGP 邻居传递 IPv4 路由信息。

在 IPv6 环境中的许多配置和 IPv4 环境下的配置都相同，只不过是用 IPv6 地址表达而已。比如对等体组、路由反射器，以及使用 route-map 调整 BGP 属性值以期达到调控路由的目的。

第 9 章　IPv6 的过渡技术

在全网还没有实现 IPv6 的现网中，如果需要把两个或多个由 IPv4 网络隔离的 IPv6 网络连接起来，使用的方式就是隧道——把整个 IPv6 报文作为 IPv4 报文的数据部分，封装在 IPv4 数据包中，穿越 IPv4 网络到达对端的 IPv6 网络，如图 9-1 所示。这种方式称为覆盖模式隧道（Overlay Tunneling）。

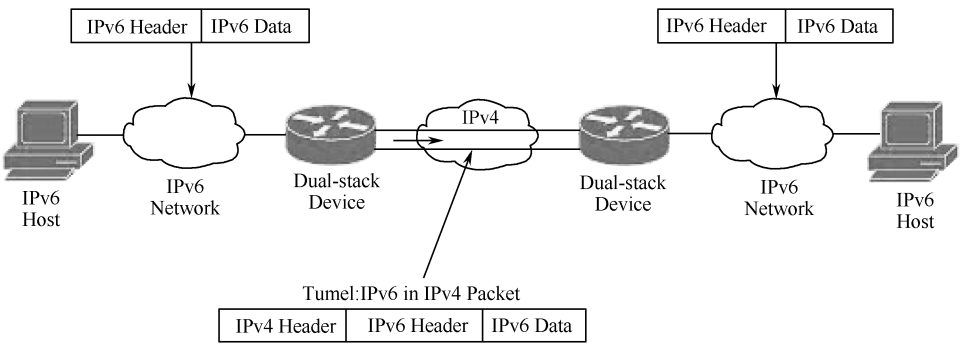

图 9-1　覆盖隧道模式示意图

在这种模式下，位于两个系统边缘的设备必须运行双协议栈（Dual Protocol Stack），即同时运行 IPv4 和 IPv6 路由协议。

9.1　隧道的实现模式

使用覆盖方法把 IPv6 的报文封装在 IPv4 报文里，以到达穿越 IPv4 网络的目的，具体的实现模式包括以下几种。

- 手动隧道（Manual）；
- 自动 6to4 隧道 （Automatic 6to4）；
- 通用路由封装（Generic Routing Encapsulating, GRE）隧道；
- ISATAP 隧道（Intrasite Automatic Tunnel Addressing Protocol, ISATAP）；
- IPv4 兼容隧道（IPv4-compatible）；
- 6RD 隧道 （IPv6 Rapid Deployment）。

根据使用环境的不同，在部署隧道时，需要考虑它们的特点。

- 手动隧道是简单的点到点隧道：仅具备携带 IPv6 数据包的能力，可以部署在多个站点间，也可以在站点内部署。
- GRE 隧道是点到点隧道：不仅可以携带 IPv6 数据包，还可以携带其他协议数据包，如 CLNS 等。可以部署在多个站点间，也可以在站点内部署。
- 6to4 隧道是点到多点自动隧道：用来连接被 IPv4 隔离的多个运行 IPv6 协议的站点，

站点使用 2002::/16 的地址前缀。

- ISATAP 隧道点到多点的自动隧道模式：用在站点内的不同 IPv6 系统间的互连。
- 6RD 隧道是对 6to4 隧道的扩展：扩展的特性使得 ISP 具备在其 IPv4 的网络上为客户提供原生 IPv6 服务的能力；借助于 IPv6 前缀代理（IPv6 Prefix Delegation）技术，可以为客户分配 IPv6 地址段。这种隧道模式适合 ISP 使用。
- IPv4 兼容隧道是点到多点隧道：使用::/96 的地址前缀，该种隧道模式不常用。

9.2　手 动 隧 道

手动配置的隧道相当于穿越 IPv4 网络的两个 IPv6 路由域之间的 PVC。建立隧道的边界路由器必须双协议栈运行。为了保证稳定的连接，隧道接口静态配置 IPv6 地址，建立隧道的源接口和隧道的终结接口配置 IPv4 地址。

1. 配置命令的语法

（1）配置隧道接口和地址

```
Device(config)# interface tunnel tunnel-number
Device(config-if)# ipv6 address {ipv6-address/prefix-length | prefix- name sub-bits/prefix-length}
或者
Device(config-if)# ipv6 address ipv6-prefix/prefix-length [eui-64]
```

（2）配置隧道源

```
Device(config-if)# tunnel source {interface | ip-address}
```

可以声明隧道起源于哪个接口或接口地址。如果使用接口，该接口必须配置了地址。

（3）配置隧道终点

```
Device(config-if)# tunnel destination ip-address
```

声明隧道终结在哪个 IPv4 地址上。

（4）配置隧道模式

```
Device(config-if)# tunnel mode ipv6ip
```

2. 配置示例

下面以图 9-2 所示的拓扑为例介绍如何配置手工隧道。如图所示，路由器 V6-1 和 V6-2 只运行 IPv6 路由协议，路由器 V4/6-1 和 V4/6-2 运行双协议。跨越 IPv4 网络建立隧道，把两个分离的 IPv6 网络连起来。在 IPv6 路由器上设置的 Loopback 接口，用来模拟 IPv6 中的路由。

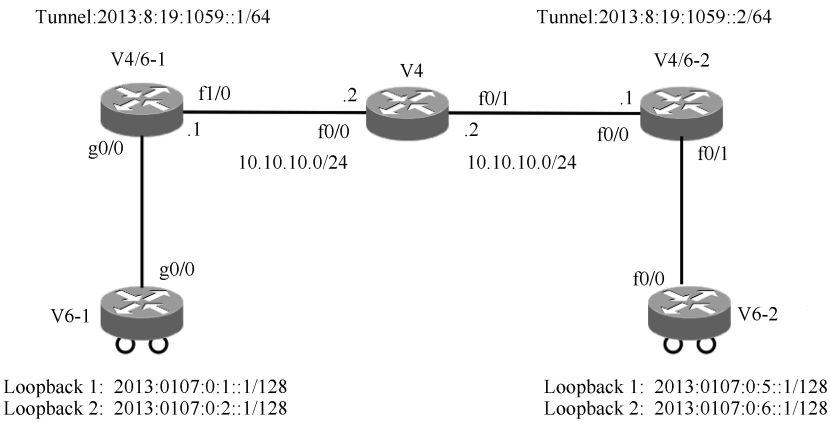

图 9-2　手动隧道示例

下面是各个路由器的配置清单。

（1）路由器 V6-1 的相关配置

```
interface Loopback1
 no ip address
 ipv6 address 2013:107:0:1::1/128
 ipv6 rip aa enable
!
interface Loopback2
 no ip address
 ipv6 address 2013:107:0:2::1/128
 ipv6 rip aa enable
!
interface GigabitEthernet0/0
 no ip address
 duplex full
 speed 1000
 media-type gbic
 negotiation auto
 ipv6 enable
 ipv6 rip aa enable
!
ipv6 router rip aa
```

（2）路由器 V4/6-1 的相关配置

```
interface Tunnel0
 no ip address
 ipv6 address 2013:8:19:1059::2/64
 ipv6 rip aa enable
```

```
 tunnel source FastEthernet1/0
 tunnel destination 10.10.11.1
 tunnel mode ipv6ip
!
interface GigabitEthernet0/0
 no ip address
 duplex full
 speed 1000
 media-type gbic
 negotiation auto
 ipv6 enable
 ipv6 rip aa enable
!
interface FastEthernet1/0
 ip address 10.10.10.1 255.255.255.0
 duplex auto
 speed auto
!
ip route 10.10.11.0 255.255.255.0 10.10.10.2
!
ipv6 router rip aa
```

配置一条静态路由的目的是保证 IPv4 网络内有路由可达路由器 V4/6-2。

（3）路由器 V4 的相关配置

```
interface FastEthernet0/0
 ip address 10.10.10.2 255.255.255.0
 duplex auto
 speed auto
!
interface FastEthernet0/1
 ip address 10.10.11.2 255.255.255.0
 duplex auto
 speed auto
```

（4）路由器 V4/6-2 的相关配置

```
interface Tunnel0
 no ip address
 ipv6 address 2013:8:19:1059::1/64
 ipv6 rip aa enable
 tunnel source FastEthernet0/0
 tunnel destination 10.10.10.1
```

```
 tunnel mode ipv6ip
!
interface FastEthernet0/1
 no ip address
 duplex auto
 speed auto
 ipv6 enable
 ipv6 rip aa enable
!
ip route 10.10.10.0 255.255.255.0 10.10.11.2
!
ipv6 router rip aa
!
```

（5）路由器 V6-2 的相关配置

```
interface Loopback1
 no ip address
 ipv6 address 2013:107:0:5::1/128
 ipv6 rip aa enable
!
interface Loopback2
 no ip address
 ipv6 address 2013:107:0:6::1/128
 ipv6 rip aa enable
!
interface FastEthernet0/0
 no ip address
 duplex auto
 speed auto
 ipv6 enable
 ipv6 rip aa enable
!
!
ipv6 router rip aa
```

3. 查看运行状况

（1）查看 V6-1 的路由表

```
V6-1#sh ipv6 route
IPv6 Routing Table - Default - 6 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
```

```
            B - BGP, HA - Home Agent, MR - Mobile Router, R - RIP
            I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
            D - EIGRP, EX - EIGRP external
            O - OSPF Intra, OI - OSPF Inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
            ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
R     2013:8:19:1059::/64 [120/2]
       via FE80::C80D:3FF:FE76:8, GigabitEthernet0/0
LC    2013:107:0:1::1/128 [0/0]
       via Loopback1, receive
LC    2013:107:0:2::1/128 [0/0]
       via Loopback2, receive
R     2013:107:0:5::1/128 [120/4]
       via FE80::C80D:3FF:FE76:8, GigabitEthernet0/0
R     2013:107:0:6::1/128 [120/4]
       via FE80::C80D:3FF:FE76:8, GigabitEthernet0/0
L     FF00::/8 [0/0]
       via Null0, receive
V6-1#
```

输出显示，已经学习到 V6-2 上的路由。

（2）跟踪去往 2013:107:0:5::1/128 的路径

```
V6-1#trace ipv6 2013:107:0:5::1

Type escape sequence to abort.
Tracing the route to 2013:107:0:5::1

  1 2013:8:19:1059::2 44 msec 20 msec 20 msec
  2 2013:8:19:1059::1 60 msec 72 msec 60 msec
  3 2013:107:0:5::1 92 msec 88 msec 84 msec
V6-1#
```

跟踪路径报告 3 跳——因为隧道是被看作点到点链路，所以是 3 跳。

（3）查看收到的路由度量值

```
V6-1#debug ipv6 rip tunnel
*Aug 20 11:04:46.787: RIPng: Packet waiting
*Aug 20 11:04:46.787: RIPng: response received from FE80::C803:6FF:FE53:8 on GigabitEthernet0/0 for aa
*Aug 20 11:04:46.787:          src=FE80::C803:6FF:FE53:8 (GigabitEthernet0/0)
*Aug 20 11:04:46.787:          dst=FF02::9
*Aug 20 11:04:46.787:          sport=521, dport=521, length=72
*Aug 20 11:04:46.787:          command=2, version=1, mbz=0, #rte=3
*Aug 20 11:04:46.787:          tag=0, metric=1, prefix=2013:8:19:1059::/64
```

```
*Aug 20 11:04:46.787:          tag=0, metric=3, prefix=2013:107:0:5::1/128
*Aug 20 11:04:46.787:          tag=0, metric=3, prefix=2013:107:0:6::1/128
```

输出显示，收到的度量值是 3 跳。在路由表里显示 4 跳，是把自己这 1 跳加上了。

（4）再看从 Loopback 接口通告的度量值

```
*Aug 20 11:04:51.191: RIPng: Sending multicast update on Loopback1 for aa
*Aug 20 11:04:51.191:          src=FE80::C804:6FF:FE53:6
*Aug 20 11:04:51.191:          dst=FF02::9 (Loopback1)
*Aug 20 11:04:51.191:          sport=521, dport=521, length=112
*Aug 20 11:04:51.191:          command=2, version=1, mbz=0, #rte=5
*Aug 20 11:04:51.191:          tag=0, metric=1, prefix=2013:107:0:1::1/128
*Aug 20 11:04:51.191:          tag=0, metric=1, prefix=2013:107:0:2::1/128
*Aug 20 11:04:51.191:          tag=0, metric=2, prefix=2013:8:19:1059::/64
*Aug 20 11:04:51.191:          tag=0, metric=4, prefix=2013:107:0:5::1/128
*Aug 20 11:04:51.191:          tag=0, metric=4, prefix=2013:107:0:6::1/128
```

（5）查看 V4/6-1 的路由表及隧道状态

```
V4/6-1#sh ipv6 route
IPv6 Routing Table - Default - 7 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
       B - BGP, HA - Home Agent, MR - Mobile Router, R - RIP
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       D - EIGRP, EX - EIGRP external
       O - OSPF Intra, OI - OSPF Inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
C    2013:8:19:1059::/64 [0/0]
      via Tunnel0, directly connected
L    2013:8:19:1059::2/128 [0/0]
      via Tunnel0, receive
R    2013:107:0:1::1/128 [120/2]
      via FE80::C80E:3FF:FE76:8, GigabitEthernet0/0
R    2013:107:0:2::1/128 [120/2]
      via FE80::C80E:3FF:FE76:8, GigabitEthernet0/0
R    2013:107:0:5::1/128 [120/3]
      via FE80::A0A:B01, Tunnel0
R    2013:107:0:6::1/128 [120/3]
      via FE80::A0A:B01, Tunnel0
L    FF00::/8 [0/0]
      via Null0, receive
V4/6-1#
V4/6-1#sh ipv6 tunnel
```

```
Tun Route    LastInp                Packets Description
  0 RIPng 00:00:26                  216
V4/6-1#
```

（6）在 V4/6-2 上查看隧道状态

```
V4/6-2#debug tunnel
Tunnel Interface debugging is on
V4/6-2#
*Aug 20 10:44:42.555: Tunnel0: IPv6/IP to classify 10.10.10.1->10.10.11.1 (tbl=0,"IPv4:Default" len=132 ttl=253 tos=0xE0) ok, oce_rc=0x0
*Aug 20 10:44:42.555: Tunnel0: IPv6/IP (PS) to decaps 10.10.10.1->10.10.11.1 (tbl=0, "default", len=132,ttl=253)
*Aug 20 10:44:42.559: Tunnel0: decapsulated IPv6/IP packet
```

输出显示，收到来自 10.10.10.1 的、携带 IPv6 数据的 IP 报文，而且也能正确解、封装。

9.3　自动 6to4 隧道

自动 6to4 隧道和手工隧道的最主要区别是：手工隧道是点到点模式，而 6to4 隧道是多点模式。自动 6to4 隧道把 IPv4 网络当作 NBMA 属性的网络对待，根据嵌入 IPv6 地址中的 IPv4 地址判断隧道的另一端。因此，它可以自动连接多个 IPv6 站点（Site）。

自动 6to4 隧道使用的地址是 2002:IPv4-address::/48，这个嵌入的 IPv4 地址就是站点路由器与 IPv4 网络相连的接口 IPv4 地址，紧随其后的 16 比特，用来在站点内划分子网使用。在边界上运行双协议的路由器，通过读取数据包的 IPv6 地址，提取出目标 IPv4 地址（即隧道的终点），建立隧道，把 IPv6 数据包封装在 IPv4 报文里，跨越 IPv4 网络传递到目标。这种隧道方式其实是基于报文的隧道。

图 9-3 描述了 3 个站点间的自动 6to4 场景。

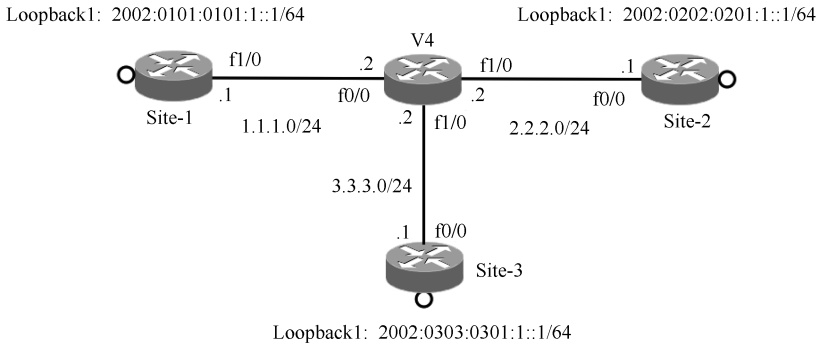

图 9-3　自动 6to4 场景

图中 Site-1、Site-2 和 Site-3 连接 IPv4 网络的接口地址分别是 1.1.1.1/24、2.2.2.1/24 和 3.3.3.1/24，把它们嵌入 IPv6 地址后，分别是 2002:0101:0101::/48、2002:0202:0201::/48 和 2002:0303:0301::/48。每个站点在这个地址段基础上利用紧随其后的 16 比特定义使用的地址段，

形成前缀为/64 的地址空间。在图中，为简化配置，分别使用了 Loopback 接口模拟业务网段。

3 台路由器的配置如下。

（1）路由器 Site-1 的配置

```
interface Loopback1
 no ip address
 ipv6 address 2002:101:101:1::1/64
!
interface Tunnel0
 no ip address
 no ip redirects
 ipv6 address 2002:101:101::1/64
 tunnel source FastEthernet1/0
 tunnel mode ipv6ip 6to4
!
interface FastEthernet1/0
 ip address 1.1.1.1 255.255.255.0
 duplex auto
 speed auto
!
ip route 2.2.2.0 255.255.255.0 1.1.1.2
ip route 3.3.3.0 255.255.255.0 1.1.1.2
!
ipv6 route 2002::/16 Tunnel0
!
```

- 配置的 IPv4 静态路由，是为了保证 IPv4 各网段之间可达；
- 配置的 IPv6 静态路由，是为了把 IPv6 数据引入隧道。

（2）路由器 Site-2 的配置

```
interface Loopback1
 no ip address
 ipv6 address 2002:202:201:1::1/64
!
interface Tunnel0
 no ip address
 no ip redirects
 ipv6 address 2002:202:201::1/64
 tunnel source FastEthernet0/0
 tunnel mode ipv6ip 6to4
!
interface FastEthernet0/0
```

```
 ip address 2.2.2.1 255.255.255.0
 duplex auto
 speed auto
!
ip route 1.1.1.0 255.255.255.0 2.2.2.2
ip route 3.3.3.0 255.255.255.0 2.2.2.2
!
ipv6 route 2002::/16 Tunnel0
!
```

（3）路由器 Site-3 的配置

```
interface Loopback1
 no ip address
 ipv6 address 2002:303:301:1::1/64
!
interface Tunnel0
 no ip address
 no ip redirects
 ipv6 address 2002:303:301::1/64
 tunnel source FastEthernet0/0
 tunnel mode ipv6ip 6to4
!
interface FastEthernet0/0
 ip address 3.3.3.1 255.255.255.0
 duplex auto
 speed auto
!
ip route 1.1.1.0 255.255.255.0 3.3.3.2
ip route 2.2.2.0 255.255.255.0 3.3.3.2
!
ipv6 route 2002::/16 Tunnel0
!
```

（4）配置完成后，检测 IPv6 的连通性

从 Site-1 上分别 Ping 另外 2 台路由器的 Loopback 接口。

```
Site-1#ping ipv6
Target IPv6 address: 2002:0202:0201:1::1
Repeat count [5]:
Datagram size [100]:
Timeout in seconds [2]:
Extended commands? [no]: y
```

```
Source address or interface: 2002:0101:0101:1::1
UDP protocol? [no]:
Verbose? [no]:
Precedence [0]:
DSCP [0]:
Include hop by hop option? [no]:
Include destination option? [no]:
Sweep range of sizes? [no]:
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 2002:202:201:1::1, timeout is 2 seconds:
Packet sent with a source address of 2002:101:101:1::1
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 28/32/44 ms
Site-1#
Site-1#ping ipv6
Target IPv6 address: 2002:0303:0301:1::1
Repeat count [5]:
Datagram size [100]:
Timeout in seconds [2]:
Extended commands? [no]: y
Source address or interface: 2002:0101:0101:1::1
UDP protocol? [no]:
Verbose? [no]:
Precedence [0]:
DSCP [0]:
Include hop by hop option? [no]:
Include destination option? [no]:
Sweep range of sizes? [no]:
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 2002:303:301:1::1, timeout is 2 seconds:
Packet sent with a source address of 2002:101:101:1::1
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 32/34/40 ms
Site-1#
```

（5）检查经由隧道的数据统计

```
Site-1#sh ipv6 tunnel
Tun Route    LastInp                    Packets Description
  0     -    00:00:06                   10
Site-1#
Site-2#sh ipv6 tunnel
```

```
Tun Route    LastInp                Packets Description
  0    -    00:00:20                    5
Site-2#
Site-3#sh ipv6 tunnel
Tun Route    LastInp                Packets Description
  0    -    00:00:16                    5
Site-3#
```

9.4　GRE over IP 隧道

使用 GRE 作为承载协议，以 IP 为传输协议，把 IPv6 作为“乘客”协议，同样可以解决跨 IPv4 网络连通 IPv6 网络的问题。把图 9-3 的描述情况，使用 GRE 隧道技术完成 IPv6 间的通信，配置清单如下。

（1）路由器 Site-1 的配置

```
interface Loopback1
 no ip address
 ipv6 address 2002:101:101:1::1/64
 ipv6 rip aa enable
!
interface Tunnel0
 no ip address
 ipv6 enable
 ipv6 rip aa enable
 tunnel source FastEthernet1/0
 tunnel destination 2.2.2.1
 tunnel mode gre ip
!
interface Tunnel1
 no ip address
 ipv6 enable
 ipv6 rip aa enable
 tunnel source FastEthernet1/0
 tunnel destination 3.3.3.1
 tunnel mode gre ip
!
!
interface FastEthernet1/0
 ip address 1.1.1.1 255.255.255.0
 duplex auto
 speed auto
```

```
!
ip route 2.2.2.0 255.255.255.0 1.1.1.2
ip route 3.3.3.0 255.255.255.0 1.1.1.2
!
!
ipv6 router rip aa
!
```

（2）路由器 Site-2 的配置

```
interface Loopback1
 no ip address
 ipv6 address 2002:202:201:1::1/64
 ipv6 rip aa enable
!
interface Tunnel0
 no ip address
 ipv6 enable
 ipv6 rip aa enable
 tunnel source FastEthernet0/0
 tunnel destination 1.1.1.1
 tunnel mode gre ip
!
interface FastEthernet0/0
 ip address 2.2.2.1 255.255.255.0
 duplex auto
 speed auto
!
ip route 1.1.1.0 255.255.255.0 2.2.2.2
ip route 3.3.3.0 255.255.255.0 2.2.2.2
!
!
ipv6 router rip aa
!
```

（3）路由器 Site-3 的配置

```
interface Loopback1
 no ip address
 ipv6 address 2002:303:301:1::1/64
 ipv6 rip aa enable
!
interface Tunnel0
```

```
 no ip address
 ipv6 enable
 ipv6 rip aa enable
 tunnel source FastEthernet0/0
 tunnel destination 1.1.1.1
 tunnel mode gre ip
!
interface FastEthernet0/0
 ip address 3.3.3.1 255.255.255.0
 duplex auto
 speed auto
!
ip route 1.1.1.0 255.255.255.0 3.3.3.2
ip route 2.2.2.0 255.255.255.0 3.3.3.2
!
ipv6 router rip aa
```

（4）配置完成后，检查运行状态

① 查看隧道建立情况：

```
Site-1#sh interfaces tunnel 0
Tunnel0 is up, line protocol is up
  Hardware is Tunnel
  MTU 17916 bytes, BW 100 Kbit/sec, DLY 50000 usec,
      reliability 255/255, txload 1/255, rxload 1/255
  Encapsulation TUNNEL, loopback not set
  Keepalive not set
  Tunnel source 1.1.1.1 (FastEthernet1/0), destination 2.2.2.1
  Tunnel protocol/transport GRE/IP
……..

Site-1#sh interfaces tunnel 1
Tunnel1 is up, line protocol is up
  Hardware is Tunnel
  MTU 17916 bytes, BW 100 Kbit/sec, DLY 50000 usec,
      reliability 255/255, txload 1/255, rxload 1/255
  Encapsulation TUNNEL, loopback not set
  Keepalive not set
  Tunnel source 1.1.1.1 (FastEthernet1/0), destination 3.3.3.1
  Tunnel protocol/transport GRE/IP
……..
```

输出显示，2 各隧道都建立起来了。承载协议是 GRE，传输协议是 IP，即建立的是 IP 上

的 GRE 隧道。

② 查看路由表：

```
Site-3#sh ipv6 route
IPv6 Routing Table - Default - 5 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
       B - BGP, HA - Home Agent, MR - Mobile Router, R - RIP
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       D - EIGRP, EX - EIGRP external
       O - OSPF Intra, OI - OSPF Inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
R      2002:101:101:1::/64 [120/2]
       via FE80::C803:9FF:FE72:6, Tunnel0
R      2002:202:201:1::/64 [120/3]
       via FE80::C803:9FF:FE72:6, Tunnel0
C      2002:303:301:1::/64 [0/0]
       via Loopback1, directly connected
L      2002:303:301:1::1/128 [0/0]
       via Loopback1, receive
L      FF00::/8 [0/0]
       via Null0, receive
Site-3#

Site-2#sh ipv6 route rip
IPv6 Routing Table - Default - 5 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
       B - BGP, HA - Home Agent, MR - Mobile Router, R - RIP
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       D - EIGRP, EX - EIGRP external
       O - OSPF Intra, OI - OSPF Inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
R      2002:101:101:1::/64 [120/2]
       via FE80::C803:9FF:FE72:6, Tunnel0
R      2002:303:301:1::/64 [120/3]
       via FE80::C803:9FF:FE72:6, Tunnel0
Site-2#

Site-1#sh ipv6 route rip
IPv6 Routing Table - Default - 5 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
       B - BGP, HA - Home Agent, MR - Mobile Router, R - RIP
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
```

```
            D - EIGRP, EX - EIGRP external
            O - OSPF Intra, OI - OSPF Inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
            ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
R     2002:202:201:1::/64 [120/2]
       via FE80::C801:9FF:FE71:8, Tunnel0
R     2002:303:301:1::/64 [120/2]
       via FE80::C802:9FF:FE71:8, Tunnel1
Site-1#
```

提示：在 IPv6 上（即把 IPv6 作为传输协议）建立 GRE 隧道也是可行的，隧道模式是 GRE IPv6。可以把 IPv6、IPv4 或路由协议作为乘客协议。另外，mGRE 也是可行的。

9.5　自动 ISATAP 隧道

Intra-Site Automatic Tunnel Addressing Protocol（ISATAP）隧道也是一种自动隧道模式，和 6to4 自动隧道机制相似，但它主要用在站点内运行 IPv6 的设备或主机间、跨越 IPv4 的网络进行 IPv6 通信，但不代表它不可以用在站点间。

承载隧道的 IPv4 网络对于 ISATAP 来说，相当于二层是 NBMA 链路。

9.5.1　ISATAP 地址

ISATAP 使用嵌入 IPv6 地址的 IPv4 地址自动建立隧道。IPv6 ISATAP 地址格式如图 9-4 所示。最高 64 比特为 IPv6 单播地址前缀（本链路地址或全球单播地址），中间 32 比特是固定的 0000:5EFE，用以代表这是 ISATAP 地址，最后 32 比特是建立隧道的 IPv4 地址（隧道源地址）。

64比特IPv6前缀	0000:5EFE	IPv6地址

图 9-4　ISATAP 地址格式

例如，如果 IPv6 前缀是 2013::/64，隧道源端口地址是 10.173.129.8，那么 ISATAP 的地址是 2013::5EFE:0AAD:8108。

9.5.2　配置 ISATAP 隧道

使用图 9-5 所示的拓扑介绍如何配置 ISATAP 隧道。

本示例计划在隧道接口启用 OSPFv3，为了简化 OSPF 的运行，避免选举 DR/BDR，在接口上把网络类型更改为点到多点非广播（Point-to-Multipoint Non-Broadcast）。在非广播型网络上，为了使 OSPF 能建立起邻居关系，必须手工指定邻居。手工指定邻居使用本链路地址即可。

为确定本链路地址，需要生成隧道接口（Tunnel 1）并在接口上启用 IPv6，然配置了隧道的起源端口和隧道的模式，这样隧道接口就有了本链路地址。在本示例中，R-3 和 R-2 的隧道源端口都是 f0/0。

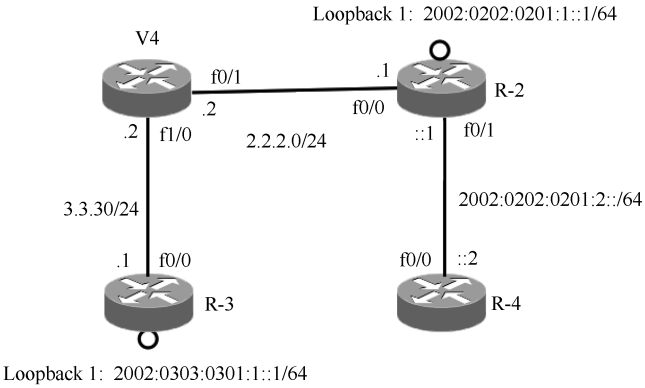

图 9-5　配置 ISATAP

例如：

```
R-2(config)# interface Tunnel1
R-2(config-if)#tunnel source FastEthernet0/0
R-2(config-if)#tunnel mode ipv6ip isatap
R-2(config-if)#ipv6 enable
```

1．查看隧道接口地址

（1）查看路由器 R-2 的隧道接口地址

```
R-2#sh ipv6 int tunnel 1
Tunnel1 is up, line protocol is up
  IPv6 is enabled, link-local address is FE80::5EFE:202:201
  No Virtual link-local address(es):
  No global unicast address is configured
  Joined group address(es):
    FF02::1
    FF02::2
    FF02::5
    FF02::1:FF02:201
  MTU is 1480 bytes
…….
```

（2）查看路由器 R-3 的隧道接口地址

```
Site-3#sh ipv6 int tunnel 1
Tunnel1 is up, line protocol is up
  IPv6 is enabled, link-local address is FE80::5EFE:303:301
  No Virtual link-local address(es):
  No global unicast address is configured
  Joined group address(es):
    FF02::1
```

```
        FF02::2
        FF02::5
        FF02::1:FF03:301
    MTU is 1480 bytes
……
```

2．路由器配置清单

（1）路由器 R-2 的配置

```
interface Loopback1
 no ip address
 ipv6 address 2002:202:201:1::1/64
 ipv6 ospf 1 area 0
!
interface Tunnel1
 no ip address
 no ip redirects
 ipv6 enable
 ipv6 ospf network point-to-multipoint non-broadcast
 ipv6 ospf neighbor FE80::5EFE:303:301
 ipv6 ospf 1 area 0
 tunnel source FastEthernet0/0
 tunnel mode ipv6ip isatap
!
interface FastEthernet0/0
 ip address 2.2.2.1 255.255.255.0
 duplex auto
 speed auto
!
interface FastEthernet0/1
 no ip address
 duplex auto
 speed auto
 ipv6 address 2002:202:201:2::1/64
 ipv6 enable
 ipv6 ospf 1 area 0
!
ip route 3.3.3.0 255.255.255.0 2.2.2.2
!
ipv6 router ospf 1
 router-id 2.2.2.2
```

```
 log-adjacency-changes
!
```

（2）路由器 R-3 的配置

```
interface Loopback1
 no ip address
 ipv6 address 2002:303:301:1::1/64
 ipv6 ospf 1 area 0
!
interface Tunnel1
 no ip address
 no ip redirects
 ipv6 enable
 ipv6 ospf network point-to-multipoint non-broadcast
 ipv6 ospf neighbor FE80::5EFE:202:201
 ipv6 ospf 1 area 0
 tunnel source FastEthernet0/0
 tunnel mode ipv6ip isatap
!
interface FastEthernet0/0
 ip address 3.3.3.1 255.255.255.0
 duplex auto
 speed auto
!
interface FastEthernet0/1
 no ip address
 shutdown
 duplex auto
 speed auto
!
ip route 2.2.2.0 255.255.255.0 3.3.3.2
!
ipv6 router ospf 1
 router-id 3.3.3.3
 log-adjacency-changes
!
```

（3）查看隧道状态

```
R-3#sh interfaces tunnel 1
Tunnel1 is up, line protocol is up
   Hardware is Tunnel
```

```
    MTU 17920 bytes, BW 100 Kbit/sec, DLY 50000 usec,
        reliability 255/255, txload 1/255, rxload 1/255
    Encapsulation TUNNEL, loopback not set
    Keepalive not set
    Tunnel source 3.3.3.1 (FastEthernet0/0)
    Tunnel protocol/transport IPv6 ISATAP
    Tunnel TTL 255
    Tunnel transport MTU 1480 bytes
…….
```

（4）查看邻居关系

```
R-3#sh ipv6 ospf nei

Neighbor ID     Pri   State           Dead Time   Interface ID    Interface
2.2.2.2           1   FULL/  -        00:01:49    8               Tunnel1
R-3#
R-2#sh ipv6 ospf neighbor

Neighbor ID     Pri   State           Dead Time   Interface ID    Interface
3.3.3.3           1   FULL/  -        00:01:55    10              Tunnel1
R-2#
```

（5）查看路由表

```
R-2#sh ipv6 route
IPv6 Routing Table - Default - 6 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
       B - BGP, HA - Home Agent, MR - Mobile Router, R - RIP
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       D - EIGRP, EX - EIGRP external
       O - OSPF Intra, OI - OSPF Inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
       ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
C    2002:202:201:1::/64 [0/0]
      via Loopback1, directly connected
L    2002:202:201:1::1/128 [0/0]
      via Loopback1, receive
C    2002:202:201:2::/64 [0/0]
      via FastEthernet0/1, directly connected
L    2002:202:201:2::1/128 [0/0]
      via FastEthernet0/1, receive
O    2002:303:301:1::1/128 [110/1000]
      via FE80::5EFE:303:301, Tunnel1
```

```
L     FF00::/8 [0/0]
        via Null0, receive
R-2#
R-3#sh ipv6 route
IPv6 Routing Table - Default - 5 entries
Codes: C - Connected, L - Local, S - Static, U - Per-user Static route
           B - BGP, HA - Home Agent, MR - Mobile Router, R - RIP
           I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
           D - EIGRP, EX - EIGRP external
           O - OSPF Intra, OI - OSPF Inter, OE1 - OSPF ext 1, OE2 - OSPF ext 2
           ON1 - OSPF NSSA ext 1, ON2 - OSPF NSSA ext 2
O     2002:202:201:1::1/128 [110/1000]
        via FE80::5EFE:202:201, Tunnel1
O     2002:202:201:2::/64 [110/1001]
        via FE80::5EFE:202:201, Tunnel1
C     2002:303:301:1::/64 [0/0]
        via Loopback1, directly connected
L     2002:303:301:1::1/128 [0/0]
        via Loopback1, receive
L     FF00::/8 [0/0]
        via Null0, receive
R-3#
```

9.6　兼容 IPv4 地址隧道

兼容 IPv4（IPv4-compatible）地址隧道使用兼容 IPv4 的 IPv6 地址自动建立，该 IPv6 地址的最高 96 比特为 0，最后 32 比特是 IPv4 地址，表达格式是 ::A.B.C.D，“A.B.C.D”代表嵌入的 IPv4 地址。

隧道的源地址就是使用建立隧道的接口的 IPv4 地址形成兼容 IPv4 的 IPv6 的地址，隧道的目标地址自动从兼容 IPv4 的 IPv6 地址中提取出来。

下面以图 9-6 所示的拓扑为例介绍这种隧道的实现。

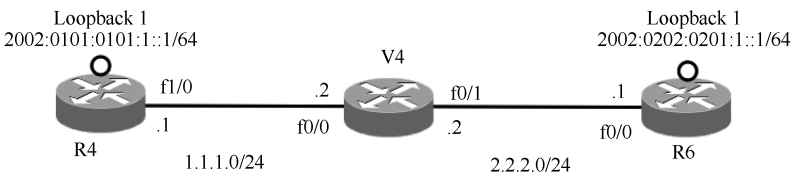

图 9-6　兼容 IPv4 地址隧道

（1）在路由器 R4 上生成一个隧道接口（Tunnel 0）

```
R4(config)#int t0
R4(config-if)#tunnel source f1/0
```

```
R4(config-if)#tunnel mode ipv6ip auto-tunnel
```

然后查看接口的相关参数：

```
R4#sh int t0
Tunnel0 is up, line protocol is up
  Hardware is Tunnel
  MTU 17920 bytes, BW 100 Kbit/sec, DLY 50000 usec,
      reliability 255/255, txload 1/255, rxload 1/255
  Encapsulation TUNNEL, loopback not set
  Keepalive not set
  Tunnel source 1.1.1.1 (FastEthernet1/0)
  Tunnel protocol/transport IPv6 auto-tunnel
  Tunnel TTL 255
  Tunnel transport MTU 1480 bytes
……

R4#sh ipv6 int t0
Tunnel0 is up, line protocol is up
  IPv6 is enabled, link-local address is FE80::101:101
  No Virtual link-local address(es):
  Global unicast address(es):
    ::1.1.1.1, subnet is ::/96
  Joined group address(es):
    FF02::1
    FF02::2
    FF02::1:FF01:101
  MTU is 1480 bytes
……..
```

输出显示，系统自动为 Tunnel 0 接口赋予了兼容 IPv4 的 IPv6 地址（::1.1.1.1）。

（2）路由器 R4 的与建立隧道相关的完整配置

```
interface Loopback1
 no ip address
 ipv6 address 2002:101:101:1::1/64
!
interface Tunnel0
 no ip address
 no ip redirects
 tunnel source FastEthernet1/0
 tunnel mode ipv6ip auto-tunnel
!
interface FastEthernet1/0
```

```
 ip address 1.1.1.1 255.255.255.0
 duplex auto
 speed auto
!
!
ip route 2.2.2.0 255.255.255.0 1.1.1.2
!
ipv6 route ::/0 ::2.2.2.1
```

使用静态路由，并使用声明下一跳的方式，指示隧道的终结点是路由器 R6。静态路由的下一跳使用的是 R6 的 Tunnel 0 接口地址（::2.2.2.1）。

（3）路由器 R6 的配置

```
interface Loopback1
 no ip address
 ipv6 address 2002:202:201:1::1/64
 ipv6 ospf 1 area 0
!
interface Tunnel0
 no ip address
 no ip redirects
 tunnel source FastEthernet0/0
 tunnel mode ipv6ip auto-tunnel

interface FastEthernet0/0
 ip address 2.2.2.1 255.255.255.0
 duplex auto
 speed auto
!
ip route 1.1.1.0 255.255.255.0 2.2.2.2
!
ipv6 route ::/0 ::1.1.1.1
```

（4）测试连通性

```
R4#ping ipv6 2002:202:201:1::1

Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 2002:202:201:1::1, timeout is 2 seconds:
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 28/35/48 ms
R4#
```

9.7　6RD 隧道

6RD 的全称是 6 Rapid Deployment, 它除了具有快速部署的含义外，还来源与它的发明者——Rémi Després，一位为数据网络作出许多贡献的法国工程师。

最初由法国的一家电信运营商（Free）在 RFC5569 中对该技术的原则和如何使用进行了描述，后来由 RFC5969 进行了更详细的定义，成为了标准。

6RD 是 6to4 的扩展，它们的主要区别是：

① 6RD 不要求使用 2002::/16 的前缀来嵌套隧道终结点的 IPv4 地址；

② 不必要把隧道终结点的 32 比特 IPv4 地址全部嵌套在 IPv6 的报文头中。

下面以图 9-7 为例介绍 6RD 的基本实现思路。

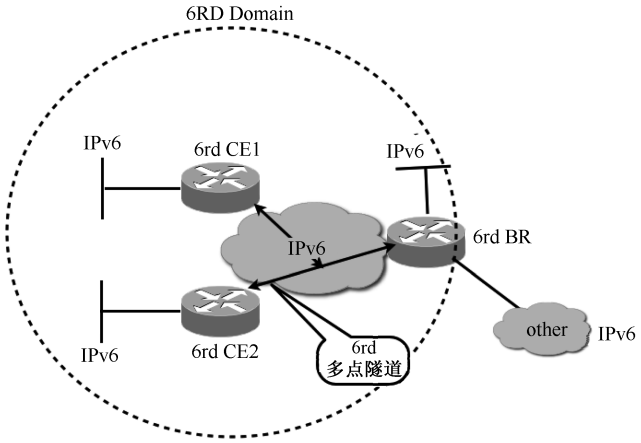

图 9-7　6RD 的实现思路

在定义的 6RD 域中，CE 和 CE 之间，以及 CE 和 BR（Border Relay）之间的 IPv6 网络通信，可以通过报文的目标 IPv6 地址中携带的部分比特，计算出隧道终结点的 IPv4 地址，从而自动建立隧道，把 IPv6 的数据封装在 IPv4 报文中，穿越 IPv4 网络到达目标 IPv6 网络。

如果 IPv6 数据的目标地址不在 6RD 域，CE 则使用 BR 的 IPv4 地址建立隧道，把数据通过隧道送给 BR，由 BR 转发到其他 IPv6 网络上。反之，来自其他 IPv6 网络、目标是 6RD 域内的某个 IPv6 网络的数据，由 BR 通过和相应的 CE 建立隧道，并把数据转发到目标网络。

9.7.1　6RD 的主要构成部件及术语

6RD 由以下两种身份的设备构成。

① Customer Edge Router：客户端边界路由器，简称 CE。它位于 ISP 的 IPv4 网络边缘，连接 IPv6 终端用户网络。IPv6 终端用户的数据由 CE 封装在 IPv4 报文中，通过隧道传送给其他 CE 或 BR。CE 路由器是双协议栈设备。

② Border Relay Router：边界中继路由器，简称 BR。它一侧通过 IPv4 连接 6RD 域内 CE 设备，另一侧连接非 6DR 域的其他 IPv6 网络。在 CE 路由器上，如果 IPv6 数据的目标地址不

在定义的 6RD 域，则需要把数据通过隧道送给 BR。BR 路由器是双协议栈设备。

在定义的 6RD 域中，还包含以下术语：

① 6RD Prefix：IPv6 地址前缀，ISP 分配给 6RD 域使用，CE 根据这个前缀判断 IPv6 数据的目标是不是在 6RD 域内。

② 6RD Delegated prefix：6RD 代理前缀，是分配给 CE 站点使用的 IPv6 地址前缀。该地址由 6RD Prefix 和建立 IPv4 隧道使用的接口 IPv4 地址比特构成，如图 9-8 所示。这里嵌入的 IPv4 地址比特可以不是 32 比特。

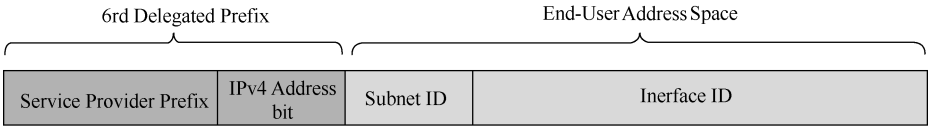

图 9-8　6RD Delegated Prefix 的构成

在该前缀后面的比特（Subnet ID 部分），可以作为 CE 所在站点的子网地址空间，用来划分多个子网使用。

③ IPv4 Common Prefix：共同 IPv4 前缀长度，指所有 CE 和 BR 隧道接口的 IPv4 地址相同的前缀部分（最高比特部分）。

④ IPv4 Common Suffix：共同 IPv4 后缀长度，指所有 CE 和 BR 隧道接口的 IPv4 地址相同的后缀部分（最低比特部分）。

例如，假设规划 6RD 域内的所有 CE 和 BR 的隧道接口地址都使用 10.1.0.0/16 地址段，并且隧道接口地址都是用 0.0.0.1，那么，它们的共同前缀就是 16 比特（10.1），共同后缀就是 8 比特（.1），剩余的 8 比特就是构成 6RD Delegated Prefix 的 IPv4 地址比特部分。

假设规划图 9-7 所示的 6RD 域使用的 6RD Prefix 是 2001:0db8::/32，那么表 9-1 列出了相关术语的具体参数值。

Parameter	Value
Service Provider Prefix	2001: DB8::/32
IPv4 Common Prefix	10.1.0.0/16
IPv4 Common Suffix	0.0.0.1/8
CE1: Delegated 6rd Prefix	2001: DB8:0100::/40
CE2: Delegated 6rd Prefix	2001: DB8:0200::/40
BR: Delegated 6rd Prefix	2001: DB8:0300::/40
CE1 (IPv4) Tunnel Transport Source	10.1.1.1
CE2 (IPv4) Tunnel Transport Source	10.1.2.1
BR (IPv4) Tunnel Transport Source	10.1.3.1

图 9-9 是在此规划下的网络部署。注意，每个站点还可以在/40 的前缀下进行子网划分。

9.7.2　隧道终结点的计算

当 IPv6 数据包到达 CE/BR 路由器，并且数据包的目标地址在 6RD 域内，CE/BR 需要把该数据包转发到相应的目标 CE/BR 路由器上，那么需要计算目标 CE/BR 路由器的 IPv4 地址，

用以建立隧道。

根据以下规则计算目标 IPv4 地址：

① 确定 IPv6 目标地址中有多少比特来自 IPv4 地址比特。

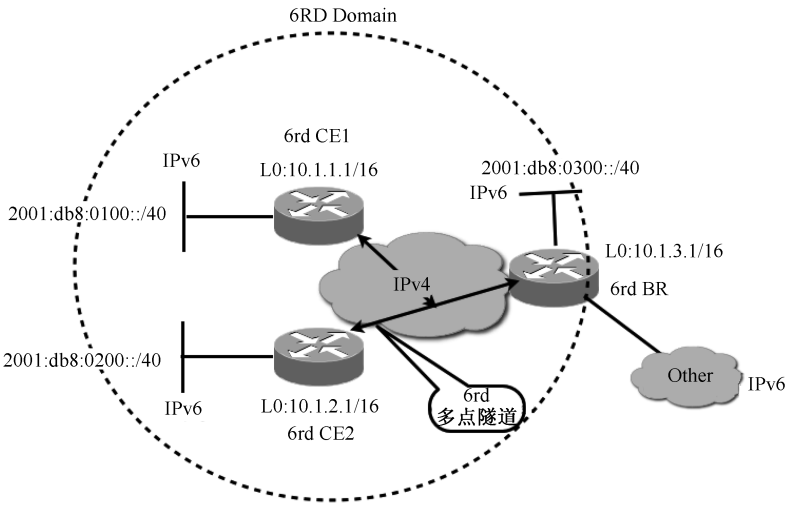

图 9-9　6RD 的部署

可以使用公式 32－（IPv4 共同前缀长度）－（IPv4 共同后缀长度）来计算。

② 确定这些比特在 IPv6 地址中的位置。它们在 6RD prefix 位置之后，如图 9-10 所示。

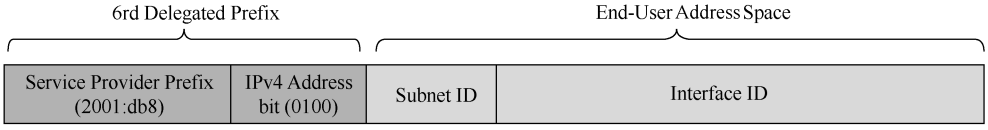

图 9-10　IPv4 地址比特位置

③ 提取出这些比特。

④ 附加上共同前缀和共同后缀部分，构成一个完整的 IPv4 地址，即为隧道终点地址，如图 9-11 所示。

图 9-11　合成 IPv4 地址

例如，假设 CE1 收到的 IPv6 报文如图 9-12 所示，他的目的地址是 2001:db8:0200::c15c:0

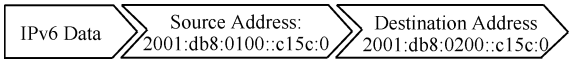

图 9-12　CE1 收到的 IPv6 报文

那么 CE1 计算隧道终点的 IPv4 的地址的方法如图 9-13 所示。

如果 CE1 收到的 IPv6 报文，目的地址不在 6RD 域内，比如 2001:db9:0200::c15c:0，那么它需要把数据通过隧道发送给 BR，即把 BR 的 IPv4 地址复制到 IPv4 报文头，如图 9-14 所示。

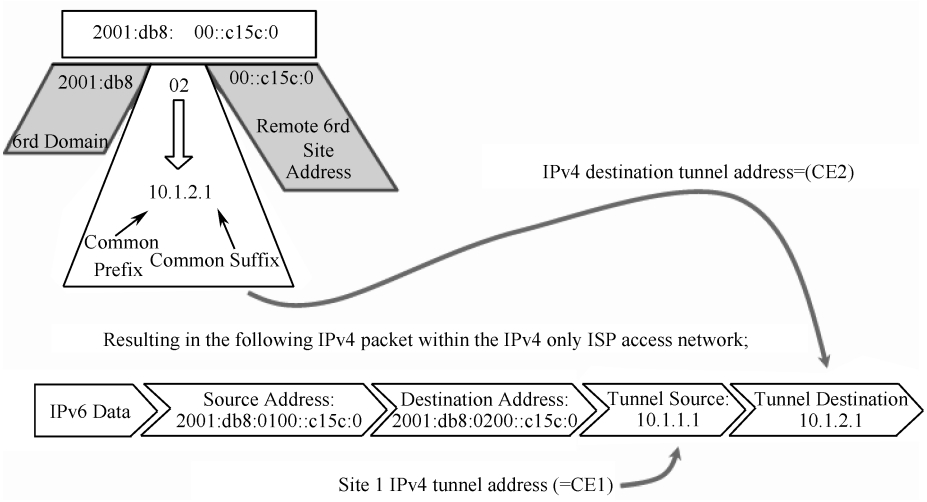

图 9-13　CE1 计算隧道终点的 IPv4 地址示意图

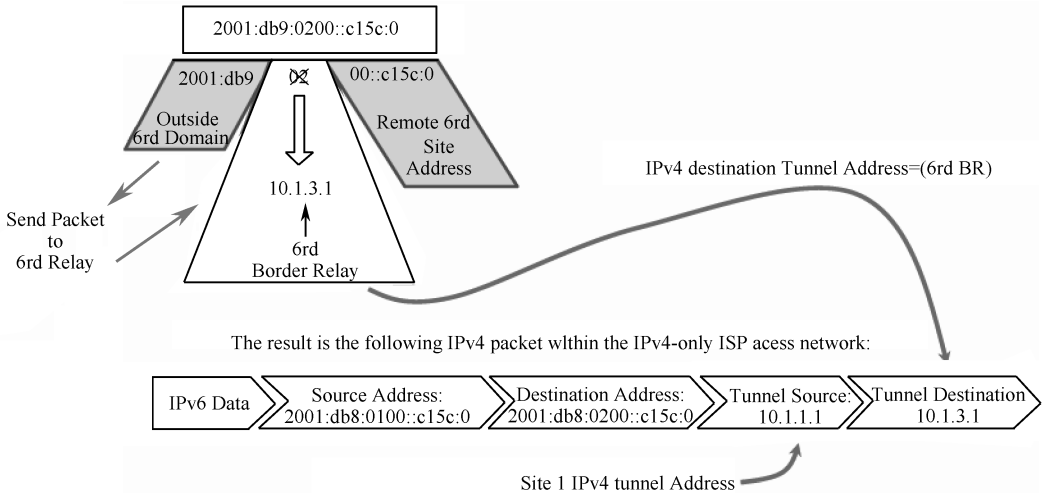

图 9-14　把 BR 地址复制到报文头

9.7.3　配置 6RD

1. 配置命令解释

在隧道接口模式下，使用如下命令配置 6RD 隧道：

```
router(config-if)# tunnel mode ipv6ip 6rd
```

该命令定义隧道模式。

```
router(config-if)# tunnel 6rd prefix {ipv6-prefix/ prefix-length}
```

该命令定义 6rd 前缀。CE 和 BR 用以判断数据的目标是不是在 6RD 域内，同时也用来判断嵌入的 IPv4 地址比特所处的位置。

```
router(config-if)# tunnel 6rd ipv4 {prefix-length } {suffix-length}
```

该命令定义共同的 IPv4 地址前缀和后缀长度，两个参数值的范围是 0～31。两个参数值的

和不能大于 31。（即最少用 1 比特代表 2 个站点）。如果不配置前缀或后缀长度，系统默认其值为 0。

```
router(config-if)# tunnel 6rd br {ipv4-address}
```

该命令定义 BR 的 IPv4 地址。

```
router(config-if)# tunnel source {ipv4-address | interface}
```

该命令定义隧道的起点 / 终点接口或接口地址。

2. 配置示例

下面以图 9-15 的拓扑为例，部署 6RD。

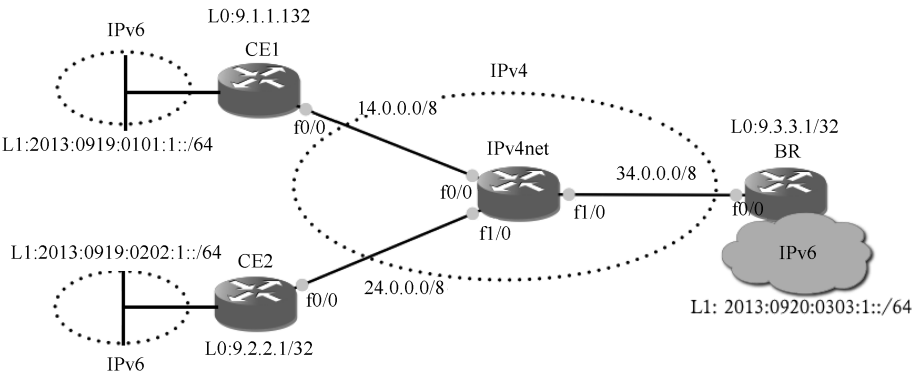

图 9-15　部署 6RD 示例

图中 CE1 和 CE2 分别表示 2 个客户站点，它们的内网使用 IPv6，为和其他 IPv6 网络互访，使用了 6RD 部署方案。路由器 BR 是 ISP 的设备，连接 IPv6 骨干网络。路由器 BR 和 CE 之间通过 IPv4 网络互通。

此次部署计划使用每台路由器的 Loopback 0 接口作为隧道始发和终结点。

规划共同 IPv4 地址前缀和后缀都是 8 比特，且接口地址都使用 1。

规划 IPv6 前缀为 2013:0919::/32

下面是路由器的配置清单。

（1）CE1 的配置清单

```
hostname CE1
!
ipv6 unicast-routing
!
interface Loopback0
 ip address 9.1.1.1 255.255.255.255
!
interface Loopback1
 no ip address
 ipv6 address 2013:919:101:1::1/64
!
```

```
interface Tunnel0
 no ip address
 no ip redirects
 ipv6 enable
 tunnel source 9.1.1.1
 tunnel mode ipv6ip 6rd
 tunnel 6rd ipv4 prefix-len 8 suffix-len 8
 tunnel 6rd prefix 2013:919::/32
 tunnel 6rd br 9.3.3.1
!
interface FastEthernet0/0
 ip address 14.0.0.2 255.0.0.0
 duplex auto
 speed auto
!
router ospf 1
 network 9.0.0.0 0.255.255.255 area 0
 network 14.0.0.0 0.255.255.255 area 0
!
!
ipv6 route ::/0 Tunnel0
```

配置 IPv6 静态默认路由，把 IPv6 数据引向隧道。

（2）路由器 CE2 的配置清单

```
hostname CE2
!
ip cef
ipv6 general-prefix 6rd 2013:919:202::/48
ipv6 unicast-routing
ipv6 cef
!
interface Loopback0
 ip address 9.2.2.1 255.255.255.255
!
interface Loopback1
 no ip address
 ipv6 address 6rd ::1:0:0:0:1/64
!
interface Tunnel0
 no ip address
 no ip redirects
```

```
 ipv6 enable
 tunnel source Loopback0
 tunnel mode ipv6ip 6rd
 tunnel 6rd ipv4 prefix-len 8 suffix-len 8
 tunnel 6rd prefix 2013:919::/32
 tunnel 6rd br 9.3.3.1
!
interface FastEthernet0/0
 ip address 24.0.0.2 255.0.0.0
 duplex auto
 speed auto
!
router ospf 1
 network 9.0.0.0 0.255.255.255 area 0
 network 24.0.0.0 0.255.255.255 area 0
!
ipv6 route ::/0 Tunnel0
!
```

（3）路由器 BR 的配置清单

```
hostname BR
!
interface Loopback0
 ip address 9.3.3.1 255.255.255.255
!
interface Loopback1
 no ip address
 ipv6 address 2013:920:303:1::1/64
!
interface Tunnel0
 no ip address
 no ip redirects
 ipv6 enable
 tunnel source Loopback0
 tunnel mode ipv6ip 6rd
 tunnel 6rd ipv4 prefix-len 8 suffix-len 8
 tunnel 6rd prefix 2013:919::/32
!
interface FastEthernet0/0
 ip address 34.0.0.2 255.0.0.0
```

```
 duplex auto
 speed auto
!
interface FastEthernet0/1
 no ip address
 duplex auto
 speed auto
ipv6 enable
!
router ospf 1
 network 9.0.0.0 0.255.255.255 area 0
 network 34.0.0.0 0.255.255.255 area 0
!
!
ipv6 route 2013:919::/32 Tunnel0
ipv6 route ::/0 FE80::C800:2FF:FE5E:7
```

（4）查看 IPv4 路由表

```
CE2#sh ip route
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      9.0.0.0/32 is subnetted, 3 subnets
O        9.1.1.1 [110/3] via 24.0.0.1, 08:49:33, FastEthernet0/0
C        9.2.2.1 is directly connected, Loopback0
O        9.3.3.1 [110/3] via 24.0.0.1, 09:14:05, FastEthernet0/0
O     14.0.0.0/8 [110/2] via 24.0.0.1, 09:14:15, FastEthernet0/0
      24.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
C        24.0.0.0/8 is directly connected, FastEthernet0/0
L        24.0.0.2/32 is directly connected, FastEthernet0/0
O     34.0.0.0/8 [110/2] via 24.0.0.1, 09:14:15, FastEthernet0/0
CE2#
```

（5）查看隧道状态

```
CE2#sh tunnel 6rd tunnel 0
Interface Tunnel0:
  Tunnel Source: 9.2.2.1
  6RD: Operational, V6 Prefix: 2013:919::/32
       V4 Prefix, Length: 8, Value: 9.0.0.0
       V4 Suffix, Length: 8, Value: 0.0.0.1
       Border Relay address: 9.3.3.1
  General Prefix: 2013:919:202::/48
CE2#
```

（6）CE1 的隧道状态

```
CE1#sh tunnel 6rd tunnel 0
Interface Tunnel0:
  Tunnel Source: 9.1.1.1
  6RD: Operational, V6 Prefix: 2013:919::/32
       V4 Prefix, Length: 8, Value: 9.0.0.0
       V4 Suffix, Length: 8, Value: 0.0.0.1
       Border Relay address: 9.3.3.1
  General Prefix: 2013:919:101::/48
CE1#
```

（7）查看某隧道端点的 6rd 代理前缀

```
CE1#sh tunnel 6rd prefix 9.2.2.1 tunnel 0
Interface: Tunnel0
Destination: 9.2.2.1
6RD Prefix: 2013:919:202::
```

输出显示，隧道地址 9.2.2.1 所代理的前缀是 2013:0919:0202::。

（8）也可以查看某 6rd 代理前缀的隧道地址

```
CE1#sh tunnel 6rd DEstination 2013:0919:0202:: tunnel 0
Interface: Tunnel0
6RD Prefix: 2013:919:202::
Destination: 9.2.2.1
CE1#
```

9.8　本 章 小 结

在没有全部迁移到 IPv6 之前，那些被 IPv4 网络互相隔离的 IPv6 网络互相通信的主要方法是采用隧道技术——把 IPv6 的数据重新封装到 IPv4 报文中，跨越 IPv4 网络通信。这种方式通

称为“覆盖”隧道模式。

本章讨论了常见的几种隧道，其中有手工配置方式，也有自动隧道方式。自动隧道包括 6to4、ISATAP、IPv4-compatible 和 6RD。

ISATAP 主要用在站点内的不同 IPv6 系统间的互连，6RD 隧道主要被运营商采纳。

参 考 文 献

[1] RFC 2080 RIPng.
[2] RFC 5340 OSPFv3.
[3] RFC 5305 IS-IS 流量工程扩展.
[4] RFC 5308 IS-ISv6.
[5] RFC 2545 利用 BGP4 多协议扩展提供 IPv6 域间路由.
[6] RFC 2547 BGP/MPLS VPNs.
[7] RFC 2858 MP-BGP 扩展.
[8] RFC 4659 BGP-MPLS IP VPN 为 IPv6 VPN 做的扩展.
[9] RFC 4798 使用 6PE 跨 IPv4 MPLS 连接 IPv6 网络.
[10] RFC 4291 IPv6 地址架构.
[11] RFC 3515 IPv6 地址结构.
[12] RFC 3587 IPv6 单播地址结构.
[13] RFC 2375 IPv6 组播地址分配.
[14] RFC 2373 IPv6 地址结构.
[15] RFC 2460 IPv6 规范.
[16] RFC 3736 DHCPv6 无状态服务.
[17] RFC 2462 IPv6 无状态地址自动配置.
[18] RFC 4862 IPv6 无状态地址自动配置.
[19] RFC 2463 ICMPv6.
[20] RFC 2464 以太网上的 IPv6 包传送.
[11] RFC 2374 可聚合全球单播地址格式.
[22] RFC 3056 跨 IPv4 的 IPv6 域互联.
[23] RFC 3068 6to4 代理路由器的任意播地址前缀.
[24] RFC 3315 DHCPv6.
[25] RFC 3633 DHCPv6 前缀代表.
[26] RFC 4214 ISATAP.
[27] RFC 3056 跨 IPv4 的 IPv6 域互联.
[28] RFC 5969 6RD 协议规范.
[29] www.cisco.com.